AD_____

Award Number: DAMD17-96-C-6104

TITLE: Clinical and Technical Evaluation of Full Field Digital Mammography

PRINCIPAL INVESTIGATOR: John M. Lewin, M.D.

CONTRACTING ORGANIZATION: University of Colorado Health Sciences Center
Denver, Colorado 80262

REPORT DATE: October 2000

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2000	Final (30 Dec 96 - 31 Mar 00)

4. TITLE AND SUBTITLE
Clinical and Technical Evaluation of Full-Field Digital Mammography

5. FUNDING NUMBERS
DAMD17-96-C-6104

6. AUTHOR(S)
John M. Lewin, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Colorado Health Sciences Center
Denver, Colorado 80262

E-MAIL:
john.lewin@uchsc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

Purpose: To evaluate full-field digital mammography (FFDM), a new technology for breast cancer detection and compare it to standard screen-film mammography (SFM) for use in screening.

Method: An initial technical evaluation of FFDM prototypes installed at two sites was performed. A clinical study followed. For the clinical study, women over 40 presenting at either of the two institutions for bilateral or screening mammography were eligible. 6768 paired FFDM and SFM exams were performed on the same subject, usually at the same visit and usually by the same technologist. The resulting studies were independently interpreted by board-certified radiologists. For each finding, interpretation included BIRADS descriptors, assessment and recommendation and a rating of probability of malignancy on a 0-100% scale for use in ROC analysis. For each finding on which the FFDM and SFM interpretations disagreed, an analysis was done to determine the cause of the disagreement.

Results: Area under the ROC curve was larger for SFM by .06, however the difference was not statistically significant ($p>.18$). SFM detected more cancers, but that difference was not statistically significant ($p>.15$). FFDM resulted in statistically significantly fewer recalls than SFM ($p<.001$). The major reasons for differences in interpretations were fortuitous positioning of normal tissue and minor differences of opinion. For differences in cancer cases, interpretation played a major role in 7 case in which SFM detected a cancer but FFDM did not.

Conclusion: We are unable to detect a difference in clinical cancer detection performance between FFDM and SFM in a screening setting.

14. SUBJECT TERMS
Mammography

15. NUMBER OF PAGES
86

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

This report summarizes the work performed in a three-year study to evaluate full-field digital mammography (FFDM) as a screening tool for breast cancer. The goal of this project is to evaluate FFDM as a screening tool for breast cancer. The study began December 30, 1996 and continued until 12/31/99 at the University of Massachusetts Medical Center (UMMC), and until 3/31/00 at the University of Colorado Health Sciences Center (UCHSC).

The first year's work on this project was devoted to acquisition and technical evaluation of two prototype full-field digital mammography systems, comparison of low-contrast lesion detection using FFDM with that of screen-film mammography (SFM), and implementation of a clinical study comparing screen-film and FFDM in screening for breast cancer. Year 1 technical evaluation results indicated that with the current GE-FFDM prototype systems, 100 micron mode had superior low-contrast detection to 50 the micron mode. In addition, we found that for compressed breasts greater than 2-4 cm thickness, grid use gave superior low-contrast detection to non-grid systems using the same technique factors. We also found that FFDM using 100 micron pixels with a grid, and with techniques matched to those of SFM, demonstrated slightly better detection of simulated low-contrast lesions than SFM with optimized optical densities (also with a grid) across the entire spectrum of breast thicknesses and compositions.

The work performed during year two of this project included: 1) additional technical evaluation directed toward optimizing clinical technique factors on the GE full-field digital detector system, and 2) continued progress on the clinical comparison of full-field digital mammography to screen-film mammography.

Year three work included the completion of the clinical comparison of full-field digital mammography to screen-film mammography and the analysis of those data. Year three was extended by three months at UCHSC, by agreement between the institution and the USAMRMC, due to a temporary interruption of patient acquisition because of IRB issues at that institution.

BODY

Technical Evaluation of FFDM [Tasks 1-3]

To optimize technique factors for FFDM, we began by studying the relationship among low-contrast lesion detection, signal-to-noise ratios (SNR), compressed breast thickness, and digital technique factors. First, for a range of simulated breast thicknesses (2-6 cm) we varied mAs at a fixed target-filter and kVp using a contrast detail phantom that allowed quantitation of low-contrast detection and signal-to-noise ratios under each imaging condition. The contrast-detail (CD) phantom was developed previously at UCHSC and consists of a 9 by 9 array of low-contrast circular test objects milled into a D-shaped 1 cm thick section of breast equivalent material, to which additional 1 cm thick sections of D-shaped breast materials are added to give the total thicknesses of 2, 4, and 6 cm. Each row of the CD pattern contained 9 low-contrast targets at a fixed level of contrast with different object diameters ranging from 0.25 mm to 4 mm. Each column contained a fixed size object, with subject contrast ranging from 0.29% to 3.95%. In the first experiment, digital images were acquired for 2, 4, and 6 cm breast thicknesses at fixed target-filter (Mo/Mo) and kVp (25) over the full allowable range of mAs values (4 mAs to 600 mAs). Signal-to-noise ratios were measured using ROI software on a uniform portion of the phantom for each exposure. CD scores were independently determined for each exposure by four medical physicists assessing detected low-contrast objects in each image under standardized viewing conditions using a standardized scoring method previously developed. CD scores represent the area of detected objects in contrast-detail space from zero (no objects detected) to 17.34 (all 81 objects detected). Signal, SNR, and CD scores were plotted versus mAs for each breast thickness to display results. CD scores also were plotted versus SNR for all breast thicknesses.

Figure 1 shows the behavior of measured mean detector signal as a function of mAs for 2, 4, and 6 cm thick simulated compressed breasts of 50% glandular/50% fatty composition. The figure shows that for each breast thickness, measured signal is linearly proportional to mAs and that for a given mAs, signal is attenuated by approximately a factor of 4 for each 2 cm increase in simulated breast thickness.

Figure 2 shows the behavior of measured SNR as a function of mAs for 2, 4, and 6 cm thick simulated compressed breasts. The figure shows that for each breast thickness, measured SNR increases approximately as the square root of mAs, as expected for a quantum-noise limited system. For a given mAs, SNR is approximately a factor of 2 lower for each 2 cm increase in simulated breast thickness.

Figure 3 shows the behavior of contrast-detail (CD) scores as a function of mAs for each breast thickness. Error bars on CD scores extend plus and minus one standard deviation about the mean, based on four independent reader's CD scores. These results demonstrate the consistency of CD scores among the four readers. They also demonstrate the clear trend of CD scores to increase rapidly for low mAs, to display decreasing improvement as mAs is increased, and to reach a plateau at high mAs. While breast dose increases linearly with mAs, these results indicate that gains in low-contrast

detection are minimal above a certain mAs value, which depends critically on breast thickness. For example, little improvement is achieved at this target-filter and kVp for mAs values beyond 50 for 2 cm breasts, 100 for 4 cm breasts, and 160 for 6 cm breasts. Moreover, these results indicate that increasing technique (in this case increasing mAs) cannot be used to overcome some of the fundamental limitations of increased breast thickness, in particular the increase in scatter-to-primary ratio, unless the system is operated at very low CD score levels. That is, for reasonable mAs settings, lesion detection in 2 cm thick breasts is going to be superior to lesion detection in 4 cm thick breasts, and , lesion detection in 4 cm thick breasts is going to be superior to lesion detection in 6 cm thick breasts.

Figure 4 shows the dependence of measured CD scores on measured SNR in each image. This figure demonstrates that the underlying reason for lower CD scores is lower SNR, independent of compressed breast thickness. This is an important result, as it indicates that the underlying basis for low-contrast lesion detection is simply SNR, independent of breast thickness. This result has a number of ramifications. It indicates that SNR in digital images (at least with this detector system) can be used as a surrogate to low-contrast lesions detection. That is, to have adequate detection of low-contrast lesions in digital mammograms, adequate SNR is required throughout the image. A simple way to assess the adequacy of FFDM technique factors is to measure the SNR in a clinical image. If that technique achieves a certain minimum SNR, one can be assured of a given level of low-contrast detection capability. For example, optimization of technique factors may be as simple as ensuring that an adequate minimum threshold of SNR (> 40-60) is achieved.

Matching Breast Doses between SFM and FFDM for Various Techniques [Tasks 4,7]

A second phase of the optimization of FFDM technique factors is to determine the constraints under which low-contrast detection capabilities will be compared experimentally. For this task, we have chosen to match digital techniques to screen-film techniques for a given breast thickness and to compare different digital techniques under the constraint of equal breast doses for a given breast thickness. All SFM image acquisition was done on a GE-DMR mammography unit using automatic optimization of parameters (AOP) mode. Kodak Min R-2000 film was used Kodak Min R- 2000 cassettes. Films were processed on a Kodak M8 processor with Kodak chemistry and autoloading. SFM phantom images were obtained with a narrow range of background film optical densities yielding maximum low-contrast detection (1.60-1.70). To equalize breast dose to SFM techniques, we measured half-value layers (HVLs) and X-ray output values at each target-filter and kVp setting on both the screen-film DMR and the FFDM system. This, along with the accepted method of calculating average glandular breast dose [1-3], allowed us to determine the FFDM techniques (in particular, mAs values) that matched breast dose to SFM for each breast thickness (2-8 cm). In addition, we have constructed an Excel software program that automatically calculates the mAs needed at each target-filter and kVp available on the FFDM system to precisely match the average glandular dose of FFDM to that of SFM. We used the parameterization of normalized average glandular dose tables by Sobol and Wu to determine technique factors that provide equal average glandular doses for different digital techniques. [4].

Figures 5-8 present the results of matching breast average glandular doses between screen-film techniques and digital techniques for 2, 4, 6, and 8 cm thick compressed breasts, respectively. For example, in **Figure 5**, the techniques listed under DMR – Initial Technique are the technique factors selected by the screen-film DMR under AOP – Contrast Mode for a 2 cm thick 50% glandular-50% fatty breast. These techniques have been set to yield a film optical density of 1.6-1.7 for a uniform 2 cm thick tissue-equivalent breast phantom. The technique factors in the lower box are FFDM techniques that exactly match the breast dose in FFDM to that in SFM. Note that target-filtration and kVp are identical between FFDM and SFM; mAs is slightly different (17 versus 16 mAs) to compensate for the differences in HVL and output of the two units. Note that the same average gladular dose is obtained on the two modalities (38.9 mrad). The table at the right indicates the mAs values that would be required at different target-filter and kVp values on the digital system to yield the same breast dose as the screen-film system. **Figures 6-8** are similar results for the digital techniques that would exactly match breast doses from SFM at 4, 6, and 8 cm breasts, respectively.

Training [Tasks 5,6]

Training for technologists and the radiologists on the use of the system was completed prior to the clinical phase. GE Medical Systems participated in this training at no charge to either site.

Clinical Evaluation of FFDM [Tasks 8-20]

The project is designed to compare FFDM to SFM in a large group of women being screened for breast cancer. The methods of the research and the statistical analysis are described in Appendix A, which is the manuscript of a paper accepted by the journal Radiology and currently *in press*.

Results in Appendix A are for the first 4965 patients, 4945 of whom were screening patients by the most strict definition. The other 20 met the entry criteria for the study and were essentially screening patients. None of the cancers reported in this study had been identified as palpable prior to the mammographic examination. The period reported in the paper is from July 1, 1997 through 5/21/99 at UCHSC and through 3/22/99 at UMMC. After these dates, a new workstation was installed at each site. The newer workstation had superior monitors and easier to use software. Each monitor was brighter and had a resolution of 2.0k x 2.5k pixels. The software included an option to display the images with an image processing function that allowed the breast to be shown to the skin line with fewer gray levels than would be possible without processing.

Results in this report are for the entire clinical study period. This period started on 7/1/97 at both institutions and ended on 12/31/99 at UMMC and 3/31/00 at UCHSC. The extra 3 months at UCHSC were to make up for a temporary cessation of the research due to a problem with that institution's investigative review board (IRB). The problem was not related to this research project, but necessitated the halting of all research studies at the institution.

Results

6768 exams were performed on 4521 women at the 2 institutions. 2538 women participated at two screenings; 291 women participated at 3 screenings. The average age of all subjects was 55.6 years.

Findings

As noted in Appendix A, individual findings were tracked in the study. There were 2048 findings recommended for recall (BIRADS 0,4 or 5) on at least one modality in the study. Of these, 690 were called on FFDM only, 1060 on SFM only and 298 on both.

The greater number of findings on SFM is statistically significant by McNemar's Chi-square test at $p < .001$. Note that only 15% of the findings in the study were called on both the SFM and FFDM readings.

Findings are distributed by mammographic type as in **Table 1**. More calcification findings were called on SFM than FFDM. The difference is statistically significant ($p < .001$).

Recall Rate

As noted in Appendix A, recall rate is calculated based on exam, rather than finding, so as to be consistent with the established definition. 1481 exams were recalled by at least one reader. 469 of these were called only on FFDM, 679 were called only on SFM and 333 were called on both. Again a BIRADS 3 assessment (probably benign) is considered negative for recall.

The recall rate for SFM was 1012/6768 = 15.0% ; the recall rate for FFDM was 802/ 6768 = 11.9 %. The higher recall rate for SFM is statistically significant ($p < .001$).

Biopsies

183 biopsies were performed for findings in the study. 88 biopsies were on findings originally detected only on SFM, 38 were for findings originally detected only on FFDM and 57 were for findings originally detected on both modalities.

As would be expected from the fact that the SFM reading led to more recalls, SFM led to more biopsies than FFDM (145 vs 95). The difference in the proportion of biopsies recommended on each modality is statistically significant ($p < .001$).

Biopsy results are as in Table 2. Nine cancers were detected only on the FFDM reading; 16 cancers were detected only on the SFM reading and 18 cancers were detected on both readings. Again BIRADS 3, probably benign, was considered a negative. If BIRADS 3 were counted as positive, there would be one less SFM-only cancer and one additional cancer detected on both modalities. These results include all cancers proven by biopsy within one year of the mammographic examinations. Cancers are counted by finding, not exam. Two cancers were in different areas of the same breast. One was detected on both modalities and the other was detected only on SFM. In order to be counted as a mammographically detected cancer, the location of the cancer had to correlate to the location of a mammographically detected finding. The difference in the number of cancers detected by each modality is not statistically significant ($p > .01$).

High risk lesions are counted as benign. Note that 6 of the 7 high risk lesions were detected only on FFDM.

The positive biopsy rate for SFM was 34/145 = 23%; the positive biopsy rate for FFDM was 27/95 = 28%.

Long-Term Surveillance **[Tasks 17,21]**
Each subject in the study is followed for at least one year (two years if possible) after their participation to ascertain if they have developed a cancer. Surveillance may occur *passively*: 1) when the women re-enrolls in the study, 2) when she returns for a mammogram outside the study or 3) when she returns to the institution for evaluation of a breast problem or *actively* by telephone or mail contact with the patient. Additionally, in Colorado, a match has been run with the state tumor registry. The match did not yield any additional cases. Because the registry has an 18-36 month lag time in recording tumors, that match will be repeated at a later date.

Of the 6768 exams in the study, 6148 were at least 1 year ago as of 9/1/00. Of these, at least 1 year surveillance is complete on 4260. Surveillance is still in progress at both sites.

Eight interval cancers were detected by surveillance of patients with negative mammograms. These cancers all became palpable within one year of screening. As all patients have not yet been followed for a year, the final number could be higher.

Ten screening cancers (i.e., those detected more than a year after study participation), have also been detected.

Sensitivity and $PPV_{screening}$
Including all 43 cancers detected by either FSM or FFDM, as well as the 8 interval cancers, the sensitivities for SFM and FFDM were 34/51 = 67% and 27/51 = 55% , respectively. The difference in the number of cancers detected was not statistically significant ($p > .1$).

$PPV_{screening}$ is defined as the fraction of recalled exams that led to a diagnosis of breast cancer. The $PPV_{screening}$ for FFDM was 27/802 = 3.4%. The $PPV_{screening}$ for SFM was the same 34/1012 = 3.4%.

Cancers by Finding Type
Table 3 gives the cancers by finding type. A trend is demonstrated in the number of cancers detected as architectural distortion. Only 3 of the 8 cancers presenting as architectural distortion were called on FFDM. This difference is not statistically significant ($p > .05$).

Discrepancy Evaluation (also referred to as Discordancy Analysis)
For all findings recalled for additional evaluation on only one of the two modalities, the two radiologists would evaluate the FFDM and SFM images side-by-side and determine by consensus a major and minor reason for the difference in interpretations, as well as the relative conspicuity of the finding on the two modalities. Conspicuity was given on a scale of 0 to 10,

where 0 meant the finding could only be seen on FFDM and 10 meant it could only be seen on SFM. The form used for discrepancy analysis is reproduced in Appendix C. Note that the 39 reasons available are divided into three broad categories: Visibility/Conspicuity, Appearance and Interpretation.

At the time of the discrepancy analysis, the radiologists had the option of dismissing a finding (i.e., not working it up) if, after viewing both images, they could determine a benign cause for the finding or felt that there was no reasonable chance that it represented cancer. This option was rarely used.

There were 1750 discrepant findings. Table 4 gives the frequency of the most common reasons given for discrepancies for findings called only on FFDM. Table 5 gives the same data for findings called only on SFM. The most common major reasons in both tables are *fortuitous positioning* and *minor difference of opinion*. *Fortuitous positioning* refers to the overlap of normal tissue causing an apparent finding or the overlap of normal tissue and an abnormality obscuring a finding on one of the studies, without obvious positioning differences. A *minor difference of opinion* is one where one reader felt that a finding had a low probability of being cancer but recalled it for workup, while the other reader did not recall it for workup, despite it having a similar appearance on both modalities.

There were 25 discrepant cancers (9 on FFDM only and 16 on SFM only). Table 6 lists the major and minor reasons given for the 9 FFDM-only cancers. No trend is appreciated in this table. The reasons are fairly evenly divided into the three main categories of Visibility/Conspicuity, Appearance and Interpretation. Table 7 lists the reasons for the 16 SFM-only cancers. Four of the cases had Minor Difference of Opinion as the reason and seven of the 16 had an Interpretation reason as the major reason.

Table 8 shows the spread of conspicuity ratings for the discrepant cancers. All 9 of the FFDM-only cancers were judged more conspicuous on FFDM, but 5 of the SFM-only cancers were judged equally conspicuous and 1 was judged more conspicuous on FFDM. This result supports the importance of interpretation in calling these cancers.

ROC (Receiver Operating Curve) Analysis

The methodology for the Free-response ROC data collection and analysis is given in Appendix A and its references. The Alternative Free-response ROC (AFROC) method of Charkraborty (6) was used. The areas under the curve are integrated using the trapezoidal method and compared using the method of Hanley and McNeil (7). The analysis was performed both by exam and by breast. Both ways the area under the SFM curve is .06 higher than that under the FFDM curve, a difference that is not statistically significant. The ROC curves by breast and by exam are shown in Figures 10 and 11. The SFM and FFDM areas are .80 and .74 by breast and .78 and .72 by exam.

Note visually how the areas are driven by the placement of the 1% point. This is a limitation of our method, in that rarely was a patient who was not called back given an ROC value greater than 0. An improvement in the method would involve more points between 0% and 1% with specific instructions to the radiologists to rate findings that are not being called back.

Discussion

In this project, we acquired and implemented prototype FFDM systems at two sites. Technical evaluation on a contrast-detail phantom showed, as expected, that FFDM performed slightly better than SFM. Optimization of technique factors showed that, given our constraint of equal doses, FFDM performed best at the same technique factors used for SFM.

We designed our clinical study to be able to detect whether FFDM was better than or worse than SFM. Because of its known technical advantages in contrast resolution and dynamic range, many assumed that FFDM would be superior to SFM in cancer detection. We did not find that result. The trends in both cancer detection and ROC analysis in our study favored SFM. As the results were not statistically significant, we cannot be 95% certain that this difference is not due to chance variation. We did find a significant difference in false positive exams, *i.e.*, recalls, favoring FFDM for having fewer. Every reader in the study had a lower recall rate on FFDM, although the range of the difference varied widely. Even though SFM had a higher recall rate, however, it maintained the same screening PPV as FFDM, due to its higher cancer detection rate.

Analysis of the causes of the differences between the SFM and FFDM interpretations showed that, for the benign findings, most of the differences were due to small random variations in the way that normal tissue overlapped, termed by us "Fortuitous Positioning" and also minor differences of opinion between the two readers. There was no dominant reason for discrepant cancers; however, almost half (7/16) of the SFM-only cancers had a major reason relating to interpretation. These included differences of opinion, errors and workstation issues. The conspicuity rating also underscored that the problem with the FFDM misses was often not visibility of the lesion.

There were several surprising findings in the study. The low level of disagreement was quite surprising. Of the 2048 findings in the study, 1750 (85%) were discrepant between the two interpretations. Also surprising was the large number of discrepant cancers, (25/43) 58%. The overall cancer detection rate of 6.4/1000 (43/6768) not including the 8 interval cancers, was also higher than would be expected from our population.

TABLE 1
Distribution of Finding Type by Modality

Finding Type	findings detected on FFDM only (%)	findings detected on SFM only (%)	Findings detected on both (%)	**Total (%)**
Mass	247 (31.3)	367 (46.5)	175 (22.2)	**789 (100)**
Asymmetric Density	246 (39.0)	339 (53.8)	45 (7.1)	**630 (100)**
Calcification	123 (27.3)	262 (58.2)	65 (14.4)	**450 (100)**
Architectural distortion	62 (43.7)	72 (50.7)	8 (5.6)	**142 (100)**
Other / not specified	12 (32.4)	20 (54.1)	5 (13.5)	**37 (100)**
Total	**690 (33.7)**	**1060 (51.8)**	**298(14.5)**	**2048 (100)**

TABLE 2
Biopsy Results

	findings detected on FFDM only	Findings detected on SFM only	Findings detected on both
Benign	23	71	39
High risk	6[1]	1[2]	0
Malignant	9	16	18

[1]3 cases of ADH; 2 of LCIS;1 of ALH
[2]LCIS

TABLE 3
Mammographically Detected Cancers By Mammographic Lesion Type

Lesion Type	Number detected on FFDM - only	Number detected on SFM - only	Number detected on both modalities
Mass	3	3	7
Asymmetric Density	1	2	3
Calcification	3	6	7
Architectural Distortion	2	5	1
Total	**9**	**16**	**18**

TABLE 4 - Most common reasons for discrepant interpretations for findings called only on FFDM. A minor reason was not required. The general classification of the reason is given in parentheses: V = visibility/conspicuity, A = appearance, I = interpretation.

Major Reason	Number	Minor Reason	Number
Fortuitous positioning (V)	239	None	343
Minor difference of opinion (I)	139	Compression difference (V)	105
Mass margin/shape more suspicious on digital (A)	62	Fortuitous positioning (V)	58
Unsure cause of visibility/contrast difference (V)	42	Minor difference of opinion (I)	47
Modality difference (V)	35	Modality difference (V)	30
Sharpness greater on digital (V)	34	Mass margin/shape more suspicious on digital (A)	19
Error in detection on film (I)	19	Sharpness greater on digital (V)	10
More calcifications visible on digital (A)	16	Ability to magnify on digital workstation (V)	10
Error in interpretation on digital (I)	14	Unsure of cause of visibility/contrast difference (V)	9
Positioning (V)	10	Error in detection on film (I)	8
Compression difference (V)	10	Other cause for visibility difference (V)	7
Digital assessed as more changed from comparison studies (I)	9	Calcification distribution more suspicious on digital (A)	5
Major difference of opinion (I)	8	Poor use of window/level on digital workstation (I)	5
Artifact simulates finding on digital (V)	7	Positioning (V)	5
Other cause for visibility difference (V)	7		

TABLE 7 - Most common reasons for discrepant interpretations for findings called only on SFM. A minor reason was not required. The general classification of the reason is given in parentheses: V = visibility/conspicuity, A = appearance, I = interpretation.

Major Reason	Number	Minor Reason	Number
Fortuitous positioning (V)	313	None	572
Minor difference of opinion (I)	280	Compression difference (V)	104
Unsure cause of visibility/contrast difference (V)	59	Minor difference of opinion (I)	73
Modality difference (V)	49	Fortuitous positioning (V)	54
Mass margin/shape more suspicious on SFM (A)	40	Modality difference (V)	37
More calcifications visible on SFM (A)	38	Workstation suboptimal	21
Error in detection on FFDM (I)	33	Error in detection on FFDM (I)	20
Positioning (V)	31	Mass margin/shape more suspicious on SFM (A)	17
Sharpness greater on SFM (V)	28	Unsure cause of visibility/contrast diff. (V)	16
Major difference of opinion (I)	20	Calcification forms more suspicious on SFM (A)	14
Calcification forms more suspicious on SFM (A)	19	Major difference of opinion (I)	13
Compression difference (V)	18	More calcifications visible on SFM (A)	13
Error in interpretation on FFDM	13	Poor use of window/level on FFDM workstation (I)	12
SFM assessed as more changed from prior (I)	12	Error in interpretation on SFM (I)	9
Poor use of window/level on FFDM workstation (I)	11	Positioning, not due to detector size diff. (V)	8
Workstation suboptimal	11	Error in interpretation on FFDM (I)	8
Poor use of other workstation controls (I)	9	Sharpness greater on SFM (V)	7
Error in interpretation on SFM (I)	9	Other cause for visibility difference (V)	7
Artifact simulates a finding on SFM (V)	9	Poor use of other workstation controls (I)	7

TABLE 8 - Most common reasons for discrepant interpretations for cancers called only on FFDM. A minor reason was not required. The general classification of the reason is given in parentheses: V = visibility/conspicuity, A = appearance, I = interpretation.

Major Reason	Number	Minor Reason	Number
Fortuitous positioning (V)	2	Mass margin/shape more suspicious on FFDM (A)	4
Sharpness greater on FFDM (V)	1	None	2
Ability to magnify on digital workstation (V)	1	Unsure cause of visibility/contrast diff. (V)	1
More calcifications visible on FFDM (A)	1	Calcification forms more suspicious on FFDM (A)	1
Mass margin/shape more suspicious on FFDM (A)	1	Dig assessed as more changed from comparison (I)	1
Error in interpretation, digital (I)*	1		
Major difference of opinion (I)	1		
Minor difference of opinion (I)	1		

*evaluation prior to obtaining diagnosis of cancer determined to be in error on digital

TABLE 9 - Most common reasons for discrepant interpretations for cancers called only on SFM. A minor reason was not required. The general classification of the reason is given in parentheses: V = visibility/conspicuity, A = appearance, I = interpretation.

Major Reason	Number	Minor Reason	Number
Minor difference of opinion (I)	4	None	7
Unsure cause of visibility/contrast diff. (V)	2	Fortuitous positioning (V)	2
Fortuitous positioning (V)	2	Poor use of workstation controls (I)	2
Mass margin/shape more suspicious on SFM (A)	2	Other: "Denser on film" (V)	1
Sharpness greater on SFM (V)	1	More calcs visible on film (A)	1
Compression difference (V)	1	Greater change on film (A)	1
Calcification forms more suspicious on SFM (A)	1	Mass margin/shaped more suspicious on SFM (A)	1
Error in detection on digital (I)	1	Error in detection on digital (I)	1
Film assessed as more changed from comparison (I)	1		
Workstation suboptimal (I)	1		

Table 10 – Relative conspicuity of discrepant cancers. Scale is 0 to 10 where 0 is a lesion only visible on FFDM and 10 is a lesion only visible on SFM.

Conspicuity	FFDM-only Findings (#)	SFM-only Findings (#)
0	1	-
1	2	-
2	2	-
3	-	1
4	4	-
5	-	5
6	-	1
7	-	5
8	-	3
9	-	2
10	-	-

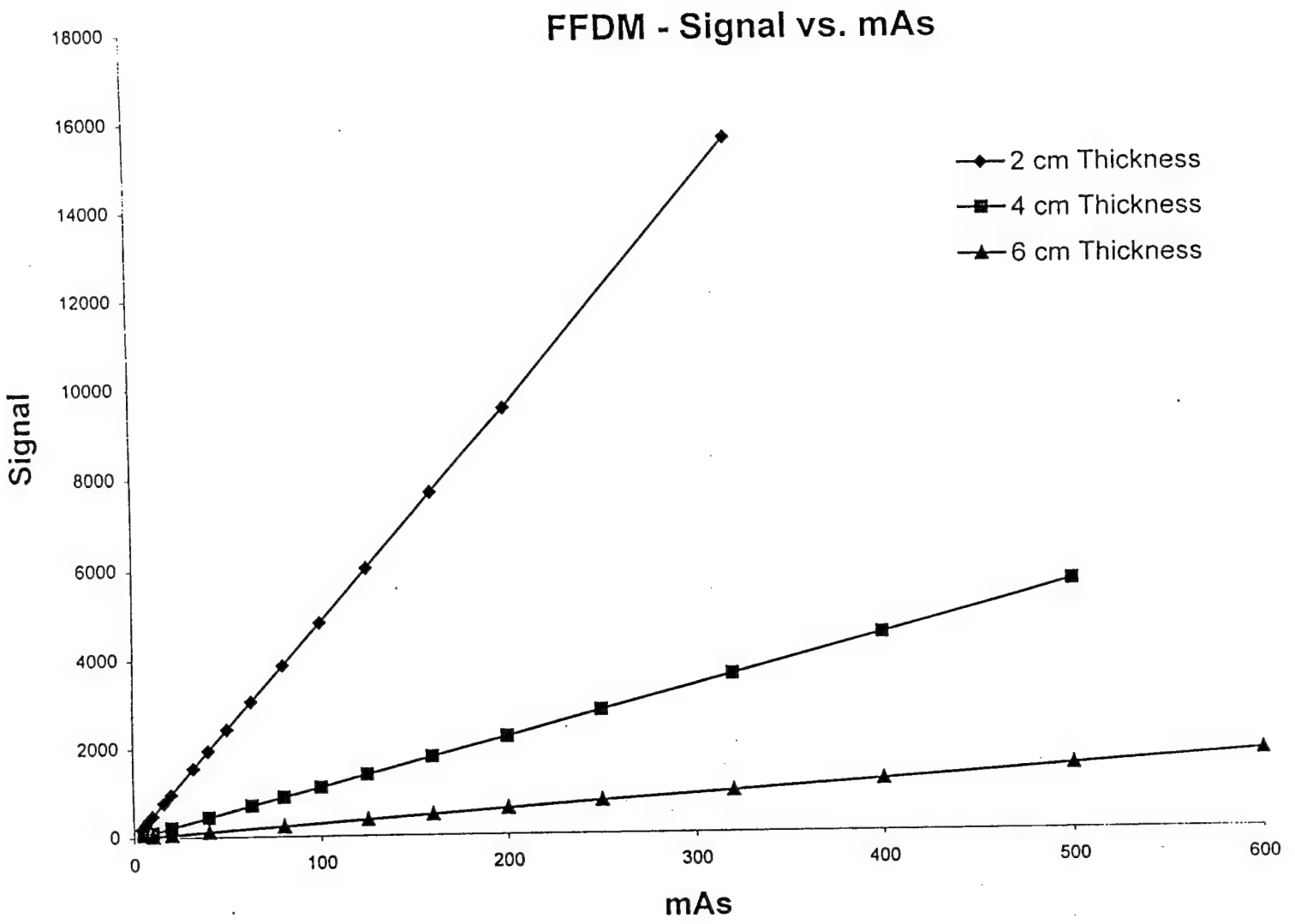

Figure 1: Detector signal measured on the GE-FFDM system versus mAs for 2, 4, and 6 cm thicknesses of 50% glandular/50% fatty compressed breasts.

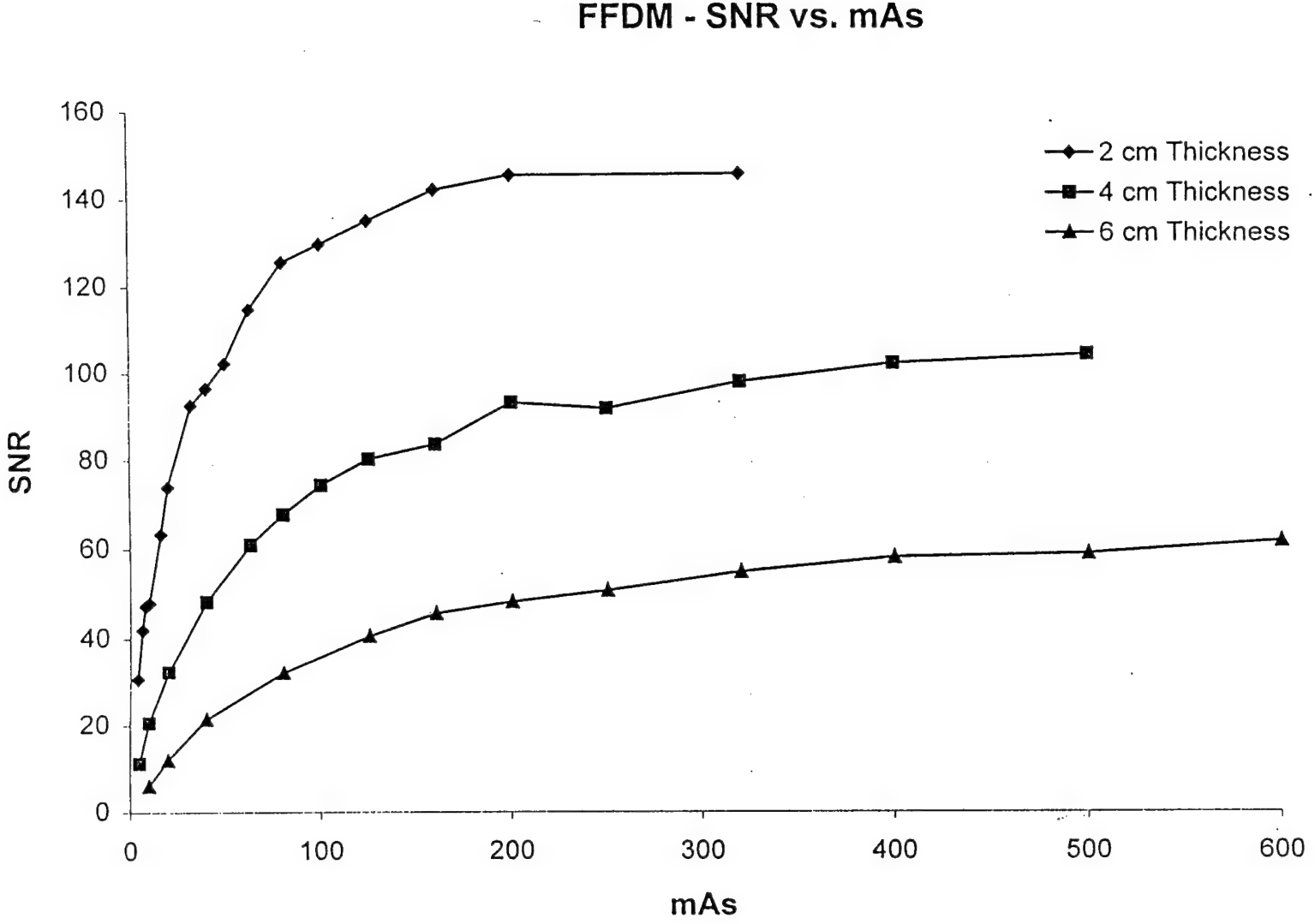

Figure 2: Signal-to-noise ratios (SNR) measured on the GE-FFDM system versus mAs for 2, 4, and 6 cm thicknesses of 50% glandular/50% fatty compressed breasts.

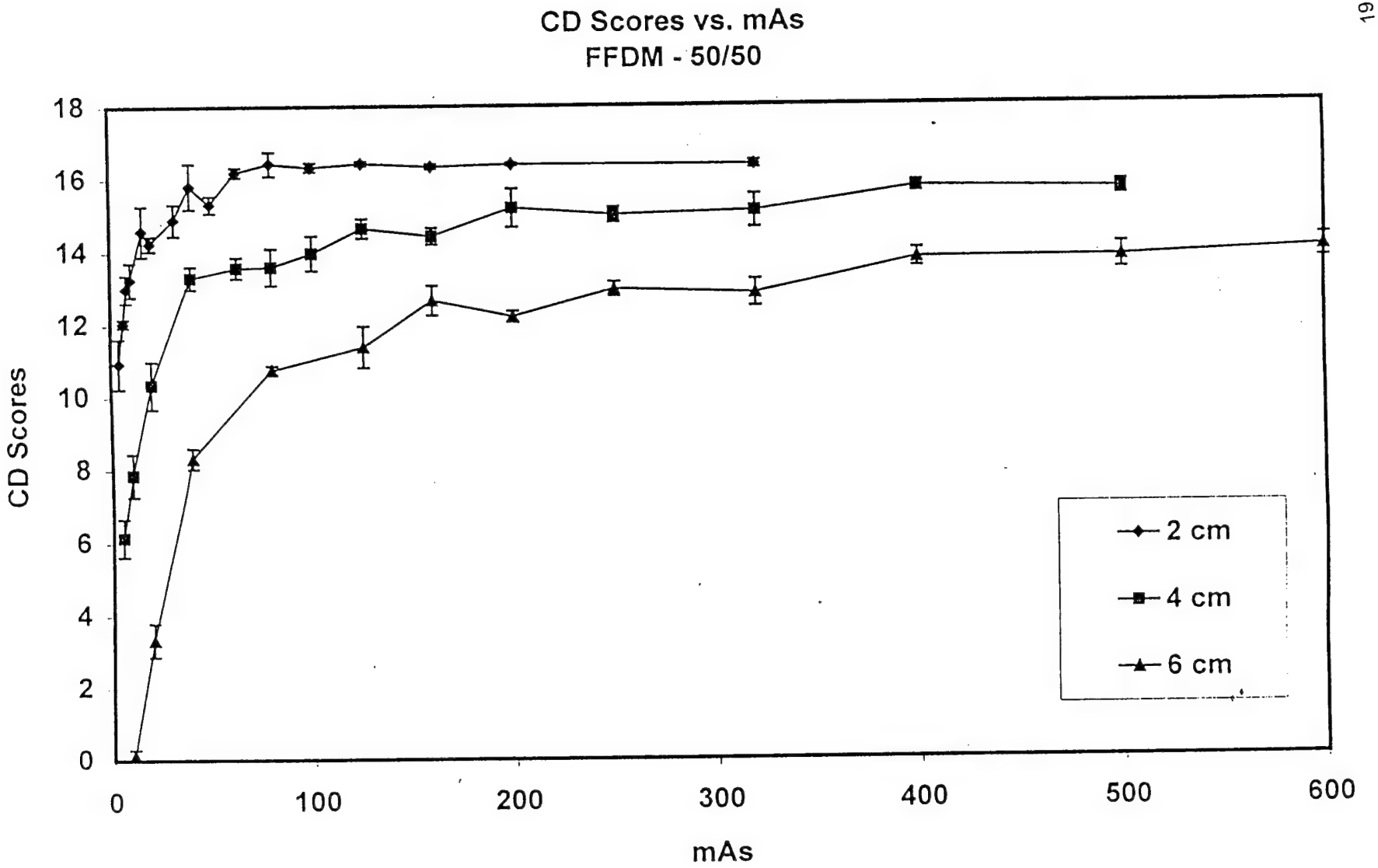

Figure 3: Contrast-detail (CD) scores measured on the GE-FFDM system versus mAs for 2, 4, and 6 cm thicknesses of 50% glandular/50% fatty compressed breasts. Point values represent mean CD scores from 4 independent readers; error bars represent ± one standard deviation among the four independent readers.

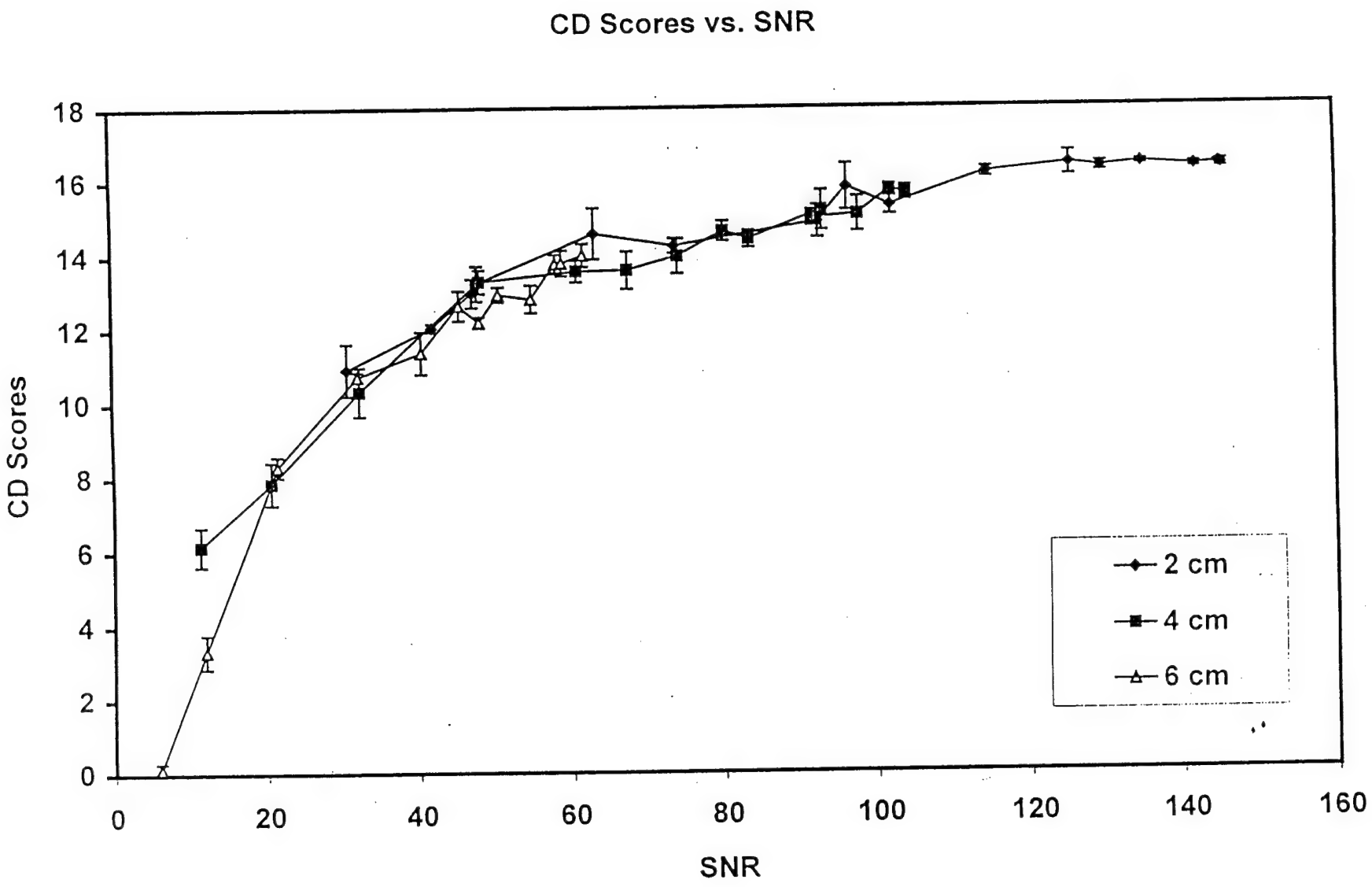

Figure 4: CD scores versus SNR for 2, 4, and 6 cm thicknesses of 50% glandular/50% fatty compressed breasts.

Mammography Technique and Dose Matching Program

Patient:	2 cm
Patient ID#:	2
Date:	1/29/99

DMR - Initial Technique

Thickness (cm)	2
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Mo/Mo
kVp	25
Comp.(100F, 50/50, 100G)	50/50
Entrance X (mR)	1245
HVL (mm Al)	0.3429
Average Glandular Dose (mrad)	38.90
Density Setting	0
mAs	16
mR/mAs	7.78
mA	100
Exposure Time (sec.)	0.16
D_{gN} (mrad/R)	312.3

DIGITAL - Matched Technique

Thickness (cm)	2
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Mo/Mo
kVp	25
Comp.(100F, 50/50, 100G)	50/50
Entrance X (mR)	1255
HVL (mm Al)	0.3397
Average Glandular Dose (mrad)	38.90
Density Setting	0
mAs	17.0
mR/mAs	7.38
mA	100
Exposure Time (sec.)	0.17
D_{gN} (mrad/R)	309.9

Digital DMR Matched Techniques

	Mo/Mo	Mo/Rh	Rh/Rh
kVp	mAs	mAs	mAs
22	30.3	35.7	
23	24.8	28.0	
24	20.3	22.6	
25	17.0	18.6	18.3
26	14.4	15.8	15.6
27	12.5	13.5	13.3
28	10.9	11.8	11.6
29	9.5	10.3	10.1
30	8.4	9.2	8.9
31	7.6	8.1	7.8
32	6.8	7.3	7.0
33	6.2	6.7	6.3
34	5.7	6.1	5.8
35	5.2	5.6	5.3
36		5.2	4.9
37		4.8	4.5
38		4.5	4.2
39		4.2	3.9
40		3.9	3.6
41			3.4
42			3.2
43			2.9
44			2.7
45			2.5
46			2.3
47			2.1
48			1.9
49			1.7

Figure 5: The upper left-hand box shows technique factors selected by the GE-DMR screen-film unit under AOP mode for a 2 cm thick 50% glandular/50% fatty compressed breast and resultant entrance exposure, half-value layer (HVL), and average glandular dose. The lower left-hand box shows technique factors for the FFDM system that match FFDM dose to SFM dose using the same target-filter and kVp. The right hand table shows mAs values that match FFDM dose to the SFM dose using different target-filter and kVp settings on the FFDM system.

Mammography Technique and Dose Matching Program

Patient:	4 cm
Patient ID#:	4
Date:	1/29/99

DMR - Intral Technique

Thickness (cm)	4
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Mo/Mo
kVp	25
Comp.(100F, 50/50, 100G)	50/50
Entrance X (mR)	705.0
HVL (mm Al)	0.3429
Average Glandular Dose (mrad)	126.38
Density Setting	0
mAs	85
mR/mAs	8.29
mA	100
Exposure Time (sec.)	0.85
D_{GN} (mrad/R)	179.3

DIGITAL - Matched Technique

Thickness (cm)	4
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Mo/Mo
kVp	25
Comp. (100F, 50/50, 100G)	50/50
Entrance X (mR)	711.0
HVL (mm Al)	0.3897
Average Glandular Dose (mrad)	126.38
Density Setting	0
mAs	85.0
mR/mAs	7.36
mA	100
Exposure Time (sec.)	0.90
D_{GN} (mrad/R)	177.8

Digital DMR Matched Techniques

	Mo/Mo	Mo/Rh	Rh/Rh
kVp	mAs	mAs	mAs
22	166.2	189.9	
23	135.0	146.9	
24	108.9	117.4	
25	90.4	96.3	94.5
26	76.1	81.0	79.3
27	65.3	68.9	66.9
28	56.7	60.1	57.2
29	48.9	52.6	49.5
30	43.4	46.7	43.0
31	39.0	41.0	37.4
32	35.0	37.2	33.3
33	31.6	33.8	29.9
34	28.8	30.7	27.3
35	26.3	28.3	24.9
36		26.2	23.0
37		24.0	21.5
38		22.4	20.0
39		20.8	18.8
40		19.4	17.6
41			16.6
42			15.4
43			14.3
44			13.5
45			12.3
46			11.4
47			10.5
48			9.6
49			8.8

Figure 6: The upper left-hand box shows technique factors selected by the GE-DMR screen-film unit under AOP mode for a 4 cm thick 50% glandular/50% fatty compressed breast and resultant entrance exposure, half-value layer (HVL), and average glandular dose. The lower left-hand box shows technique factors for the FFDM system that match FFDM dose to SFM dose using the same target-filter and kVp. The right hand table shows mAs values that match FFDM dose to the SFM dose using different target-filter and kVp settings on the FFDM system.

Mammography Technique and Dose Matching Program

Patient:	6 cm
Patient ID#:	6
Date:	1/29/99

DMR - Initial Technique

Thickness (cm)	6
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Mo/Rh
kVp	27
Comp.(100F, 50/50, 100G)	50/50
Entrance X (mR)	1576.2
HVL (mm Al)	0.2
Average Glandular Dose (mrad)	234.97
Density Setting	0
mAs	168
mR/mAs	9.38
mA	100
Exposure Time (sec.)	1.68
D_{gN} (mrad/R)	149.1

DIGITAL - Matched Technique

Thickness (cm)	6
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Mo/Rh
kVp	27
Comp.(100F, 50/50, 100G)	50/50
Entrance X (mR)	1580.5
HVL (mm Al)	0.2
Average Glandular Dose (mrad)	234.97
Density Setting	0
mAs	178.4
mR/mAs	8.86
mA	100
Exposure Time (sec.)	1.78
D_{gN} (mrad/R)	148.7

Digital DMR Matched Techniques

kVp	Mo/Mo mAs	Mo/Rh mAs	Rh/Rh mAs
22	444.7	508.7	
23	360.8	389.5	
24	290.6	308.7	
25	240.6	251.6	243.4
26	201.8	210.9	202.9
27	172.7	178.4	170.2
28	149.4	155.5	144.9
29	128.4	135.5	124.8
30	113.6	120.1	107.9
31	101.7	105.2	93.4
32	90.9	95.1	83.0
33	81.9	86.3	74.3
34	74.4	78.0	67.6
35	67.7	71.7	61.5
36		65.9	56.9
37		59.9	53.0
38		55.2	49.1
39		50.7	46.2
40		46.7	43.2
41			40.6
42			37.8
43			35.1
44			33.2
45			30.5
46			28.5
47			26.3
48			24.4
49			22.7

Figure 7: The upper left-hand box shows technique factors selected by the GE-DMR screen-film unit under AOP mode for a 6 cm thick 50% glandular/50% fatty compressed breast and resultant entrance exposure, half-value layer (HVL), and average glandular dose. The lower left-hand box shows technique factors for the FFDM system that match FFDM dose to SFM dose using the same target-filter and kVp. The right hand table shows mAs values that match FFDM dose to the SFM dose using different target-filter and kVp settings on the FFDM system.

Mammography Technique and Dose Matching Program

Patient:	8 cm
Patient ID#:	8
Date:	1/29/99

DMR - Initial Technique

Thickness (cm)	8
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Rh/Rh
kVp	28
Comp.(100F, 50/50, 100G)	50/50
Entrance X (mR)	3140.5
HVL (mm Al)	0.4315
Average Glandular Dose (mrad)	384.77
Density Setting	0
mAs	283
mR/mAs	11.10
mA	80
Exposure Time (sec.)	3.5375
D_{GN} (mrad/R)	122.5

DIGITAL - Matched Technique

Thickness (cm)	8
T/F (Mo/Mo, Mo/Rh, Rh/Rh)	Rh/Rh
kVp	28
Comp.(100F, 50/50, 100G)	50/50
Entrance X (mR)	3166.8
HVL (mm Al)	0.4274
Average Glandular Dose (mrad)	384.77
Density Setting	0
mAs	284.9
mR/mAs	11.74
mA	80
Exposure Time (sec.)	3.69
D_{GN} (mrad/R)	121.5

Digital DMR Matched Techniques

kVp	Mo/Mo mAs	Mo/Rh mAs	Rh/Rh mAs
22	920.8	1062.0	
23	745.5	807.5	
24	599.9	636.9	
25	496.5	517.5	500.3
26	416.1	432.8	414.9
27	355.7	365.9	347.0
28	307.3	318.7	294.9
29	263.8	277.8	253.8
30	233.0	246.3	219.1
31	208.4	215.8	189.6
32	186.0	195.1	168.3
33	167.3	177.0	150.4
34	151.9	159.8	136.4
35	138.1	146.6	123.7
36		134.5	114.1
37		121.6	105.8
38		111.6	97.8
39		101.6	91.5
40		92.7	85.3
41			79.9
42			74.3
43			69.0
44			65.3
45			60.1
46			56.2
47			52.1
48			48.4
49			45.1

Figure 8: The upper left-hand box shows technique factors selected by the GE-DMR screen-film unit under AOP mode for a 8 cm thick 50% glandular/50% fatty compressed breast and resultant entrance exposure, half-value layer (HVL), and average glandular dose. The lower left-hand box shows technique factors for the FFDM system that match FFDM dose to SFM dose using the same target-filter and kVp. The right hand table shows mAs values that match FFDM dose to the SFM dose using different target-filter and kVp settings on the FFDM system.

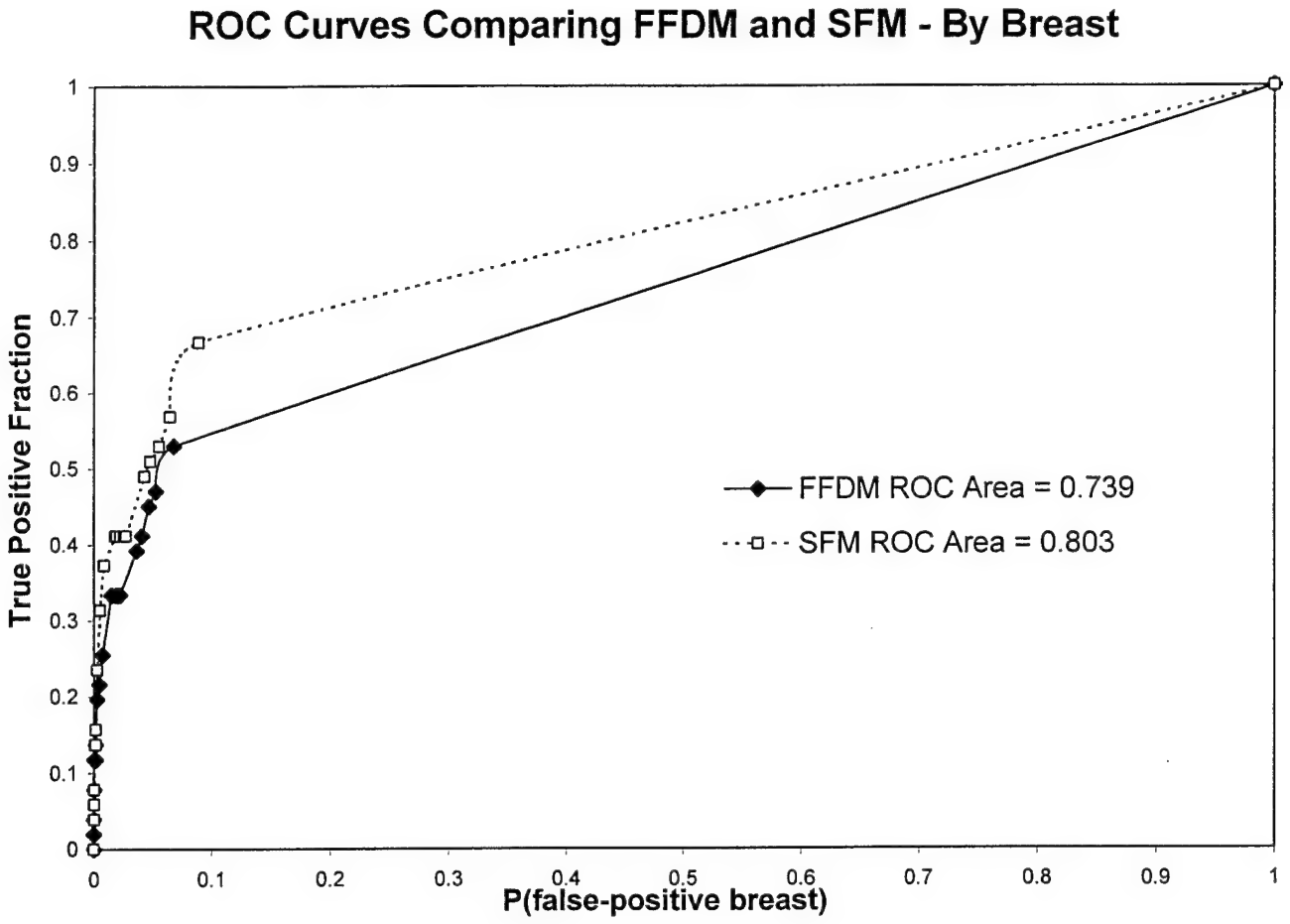

Figure 9. Alternative free-response ROC (AFROC) curves for SFM and FFDM plotted using a rating scale from 0-100. The x-axis scale is the probability of a false positive finding being called on the two screening views of a given breast. This is analogous to the false positive rate in a standard ROC experiment. The area under the SFM curve is .80; the area under the FFDM curve is .74. the difference is not statistically significant (p= .18).

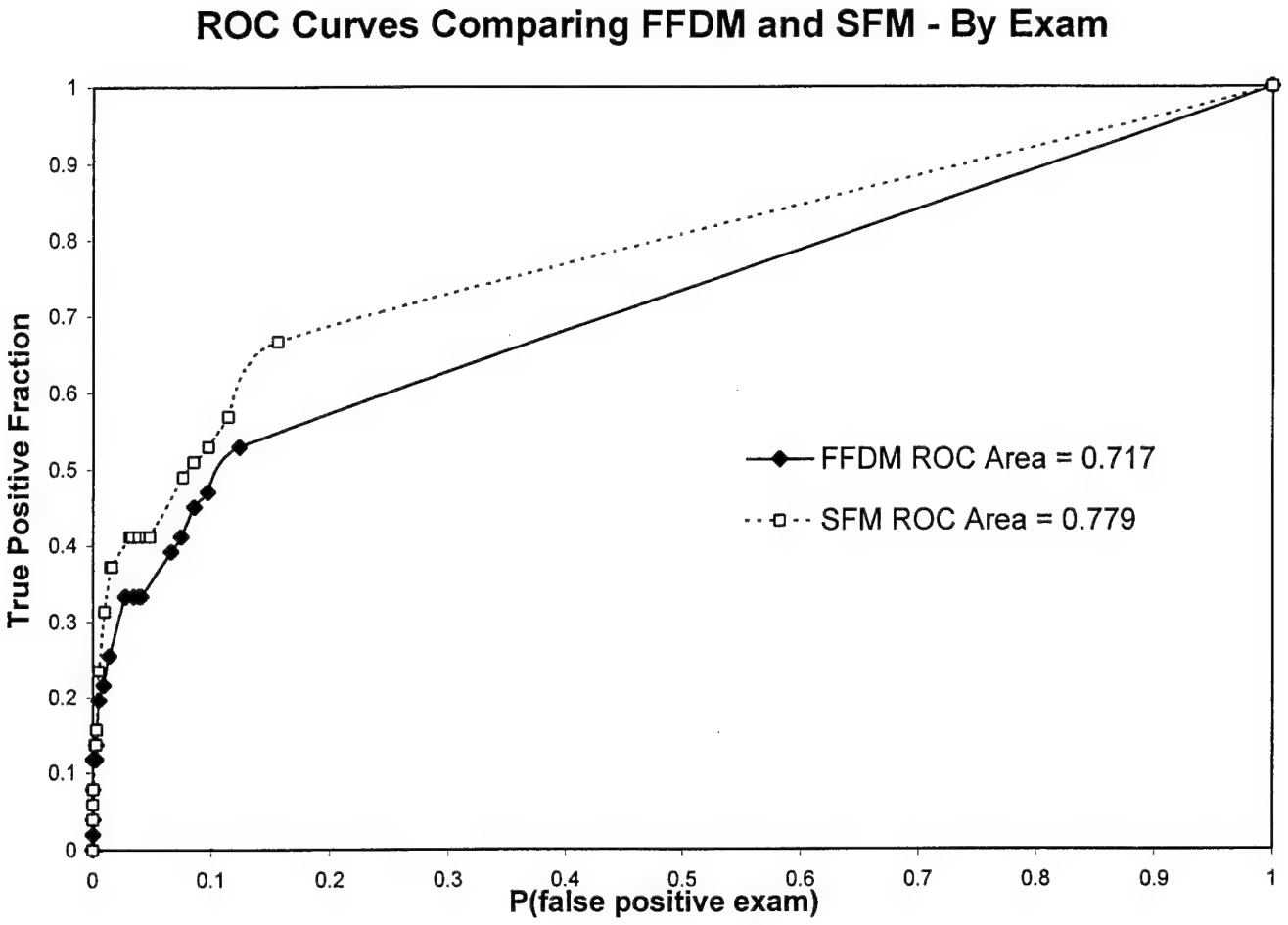

Figure 10. Alternative free-response ROC (AFROC) curves for SFM and FFDM plotted using a rating scale from 0-100. The x-axis scale is the probability of a false positive finding being called on the screening exam of a given patient. This is analogous to the false positive rate in a standard ROC experiment. The area under the SFM curve is .78; the area under the FFDM curve is .72. the difference is not statistically significant (p= .19).

KEY RESEARCH ACCOMPLISHMENTS

- Implementation of 2nd and 3rd flat panel FFDM prototype systems in the world
- Complete technical evaluation of these systems
- Only clinical research trial in the world testing FFDM against SFM for screening
  - Found no statistically significant difference in cancer detection or area under the ROC curve, with a trend favoring SFM
  - Found a statistically significant difference in recall rate, with FFDM having a lower recall rate
- Analyzed reasons for differences in interpretations between FFDM and SFM in the study

REPORTABLE OUTCOMES

Abstracts/Presentations

Hendrick RE, Berns E, Chorbajian B, Choi K, Lewin JM, Sisney GS. Low contrast lesion detection:comparison of screen-film and full-field digital mammography. Radiology 205(P) :274, 1997.

Lewin JM, Hendrick RE, D'Orsi CJ, Moss LJ, Sisney GA, et al. Clinical evaluation of a full-field digital mammography prototype for cancer detection in a screening setting-- work in progress. Presented at the 1998 RSNA Scientific Meeting. Radiology 209(P): 238, 1998.

Lewin JM, Hendrick RE, D'Orsi CJ, et al. Clinical evaluation of full-field digital mammography in a screening population. Presented at the 2000 RSNA Scientific Meeting. Radiology 217(P): 199, 2000.

Lewin JM, D'Orsi CJ, Hendrick RE, et al. Causes of differences between full-field digital mammography and screen-film mammography interpretations in the Colorado/Massachusetts full-field digital mammography screening trial. Presented at the 2000 RSNA Scientific Meeting. Radiology 217(P): 199, 2000.

Publications

Lewin JM. Digital mammography, a candid assessment. Diagnostic Imaging, September 1999, pp. 40-45.

Vedantham S, Karellas A, Suryanarayanan S, Albagli D, Han S, Tkaczyk EJ, Landberg CE, Opsahl-Ong B, Granfors PR, Levis I, D'Orsi CJ, Hendrick RE. Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype. Med Phys. 2000 Mar;27(3):558-67.

Vedantham S, Karellas A, Suryanarayanan S, D'Orsi CJ, Hendrick RE. Breast imaging using an amorphous silicon-based full-field digital mammographic system: stability of a clinical prototype. J. Digital Imaging 2000 nov;13(4):191-199.

Lewin JM, Hendrick RE, D'Orsi CJ, et al. Comparison of full-field digital mammography to screen-film mammography for cancer detection: results of 4945 paired examinations. Radiology. *In Press.*

CONCLUSIONS

1. Full-Field Digital Mammography (FFDM) was slightly technically superior to Screen-Film Mammography (SFM) in a test of low-detection objects using a contrast-detail phantom.
2. FFDM was not found to be either superior or inferior to SFM for screening, based on the area under the ROC curve and the cancer detection rate.
3. FFDM may result in fewer women being recalled for additional workup at screening.
4. There is a large amount of variability in interpretation in a paired study such as this one. Fortuitous overlap of normal structures and minor differences of opinion account for the majority of the variability.

Implications:
It had been assumed that FFDM, because of its superior performance characteristics in terms of contrast resolution and dynamic range, would detect more cancers than SFM. We showed that, although it did perform slightly better on phantoms, FFDM did not detect more cancers than did SFM and did not perform better overall than SFM. This shows that FFDM cannot be assumed to be better and new systems must be tested clinically when they are introduced. A major concern about FFDM was that it would lead to more recalls. We showed that not to be true. In our study, in fact, FFDM led to fewer recalls.

There are currently plans to conduct a very large multi-center trial modeled on this trial. Our results show that a large amount of variability can be expected and must be accounted for when powering the trial. A larger than expected number of cancers may be discovered during the trial, however.

REFERENCES

Cited References:

1. Rothenberg LN, ed. NCRP Report No. 85: Mammography: A User's Guide, Bethesda, MD: National Council on Radiation Protection and Measurement, 1986.

2. Wu X, Barnes GT, Tucker DM. Spectral dependence of glandular tissue dose in screen-film mammography. Radiology 1991; 179:143-148.

3. Wu X, Gingold EL, Barnes GT, Tucker DM. Normalized average glandular dose in molybdenum target-rhodium filter and molybdenum target-rhodium filter mammography. Radiology 1994; 193:83-89.

4. Sobol WT, Wu X. Parametrization of mammography normalized average glandular dose tables. Med. Phys. 1997; 24: 547-554.

5. Lewin JM, Hendrick RE, D'Orsi CJ, Moss LJ, Sisney GA, Karellas A. Clinical evaluation of a full-field digital mammography prototype for cancer detection in a screening setting – work in progress. Radiology 209(P): 238, 1998.

6. Chakrabotrty DP, Winter LH. Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology. 1990 Mar;174(3 Pt 1):873-81.

7. Hanley JA and McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983 Sep;148(3):839-43.

Additional References (not cited by number in this report):

Feig SA, Yaffe MJ. Current status of digital mammography. Semin. Ultrasound, CT, MR 1996; 17: 424-443.

Schmidt RA, Nishikawa RM. Clinical use of digital mammography: the present and the prospects. J. Digit. Imaging 1995; 8 (Supplement 1): 74-79.

Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. JNCI 1993; 85: 1644-56.

Elwood JM, Cox B, Richardson AK . The effectiveness of breast cancer screening by mammography in younger women. Online Journal Curr. Clin. Trials 1993; 1: 32-227.

Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992; 184: 613-617.

Jackson VP, Hendrick RE, Feig SA, Kopans DB. Imaging of the radiographically dense

breast. Radiology 1993; 188: 297-301.

Bassett LW, Gold RH, Kimme-Smith C. History of the technical development of mammography, in Syllabus: A Categorical Course in Physics: Technical Aspects of Breast Imaging, 3rd Edition, Eds. AG Haus and MJ Yaffe, 1994: 9-20.

Johns PC, Yaffe MJ. X-ray characterization of normal and neoplastic breast tissues. Phys. Med. Biol. 1987; 32: 675-695.

Cohen G, McDaniel DL, Wagner LK. Analysis of variations in contrast-detail experiments. Medical Physics 1984; 11: 469- 473.

Robson KJ, Kotre CJ, Faulkner K. The use of a contrast-detail test object in the optimization of optical density in mammography. The British Journal of Radiology 1995; 68: 277- 282.

Law J. Improved image quality for dense breasts in mammography. The British Journal of Radiology 1992; 65: 50-55.

Hendrick RE, Bassett LW, Butler PA, et. al., Mammography Quality Control: Radiologist's Manual, Radiologic Technologist's Manual, Medical Physicist's Manual, American College of Radiology, Revised Edition, 1994.

PERSONNEL WHO RECEIVED SUPPORT FROM THIS RESEARCH EFFORT

UCHSC

John M Lewin, M.D.
Pamela K. Isaacs, D.O.
RE Hendrick, Ph.D.
Virginia Vance, R.N.
Melody Lisne, R.T.

UMMC

Carl J D'Orsi, M.D.
Larry Moss, M.D.
Andrew Karellas, Ph.D.
Mary Heck, R.T.
Vanessa Brown, M.S.

APPENDICES:

A. Manuscript of publication *in press*:

Lewin JM, Hendrick RE, D'Orsi CJ, et al. Comparison of full-field digital mammography to screen-film mammography for cancer detection: results of 4945 paired examinations. Radiology. *In Press.*

B. Form used for Discrepancy (Discordancy) Evaluation

C. Reprint of article resulting from this work:

Vedantham S, Karellas A, Suryanarayanan S, Albagli D, Han S, Tkaczyk EJ, Landberg CE, Opsahl-Ong B, Granfors PR, Levis I, D'Orsi CJ, Hendrick RE. Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype. Med Phys. 2000 Mar;27(3):558-67.

D. Reprint of article resulting from this work:

Vedantham S, Karellas A, Suryanarayanan S, D'Orsi CJ, Hendrick RE. Breast imaging using an amorphous silicon-based full-field digital mammographic system: stability of a clinical prototype. J. Digital Imaging 2000 nov;13(4):191-199.

Comparison of full-field digital mammography to screen-film mammography for cancer detection: results of 4,945 paired examinations

John M. Lewin, M.D.[1], R. Edward Hendrick, Ph.D.[1,3], Carl J. D'Orsi, M.D.[2], Pamela Isaacs, D.O.[1], Lawrence J. Moss, M.D.[2], Andrew Karellas, Ph.D.[2], Gale A. Sisney, M.D.[1,4], Christopher K. Kuni, M.D.[1], Gary R. Cutter, Ph.D.[5]

[1]Department of Radiology
University of Colorado Health Sciences Center
Denver CO 80262

[2]Department of Radiology
University of Massachusetts Medical Center
Worcester MA

[3]Current address: Northwestern University Medical School
Chicago IL

[4]Current address: Georgetown University Medical School
Washington DC

[5]AMC Cancer Research Center
Denver CO

Address all correspondence to:

John M. Lewin, M.D.
Department of Radiology
University of Colorado Health Sciences Center
4200 East Ninth Avenue
Denver, Colorado 80262
Telephone: (303) 372-1627
Facsimile: (303) 372-1649
E-mail: John.Lewin@UCHSC.edu

This work was supported by the US Army Breast Cancer Research and Materiel Command, Grant DAMD17-96-C-6104, R. E. Hendrick, PI, University of Colorado Health Sciences Center.

A subset of these data were presented at the 1998 RSNA meeting.

Submitted as an **Original Research** manuscript.

Abstract

Purpose: To compare full-field digital mammography (FFDM) to screen-film mammography (SFM) for cancer detection in a screening population.

Materials and Methods: 4,945 FFDM examinations were performed on women age 40 and over presenting for SFM at two institutions. Two views of each breast were acquired with each modality. SFM and FFDM were each interpreted independently. Findings detected on either SFM or FFDM were evaluated with additional imaging and, if warranted, biopsy.

Results: 152 biopsies were performed on patients in the study, resulting in the diagnosis of 35 breast cancers. SFM detected 22 of these cancers and FFDM detected 21. Four were interval cancers that became palpable within one year of screening and were considered false negatives on both modalities. The difference in cancer detection rate was not statistically significant. FFDM had a statistically significantly lower recall rate (11.5%) than SFM (13.9%). The positive biopsy rate for findings detected by FFDM (30.4%) was higher than for findings detected by SFM (19.3%) but this difference was not statistically significant.

Conclusion: An unbiased prospective study comparing FFDM to SFM in a screening population is being conducted at two sites. No difference in cancer detection rate has yet been observed between FFDM and SFM. FFDM has so far led to fewer recalls than SFM.

Key Words: breast radiography, technology, comparative studies, digital radiography

Introduction

In recent years, advances in screen-film technology and film processing techniques have contributed to significant improvements in mammographic image quality. While screen-film mammography provides a powerful tool for detection and follow-up of suspicious lesions, it has significant limitations in detecting very subtle soft tissue lesions, especially in the presence of dense glandular tissues. One of the limitations of screen-film mammography is that the film serves simultaneously as the image receptor, display medium, and long term storage medium. This limitation can lead to loss of image contrast, especially when exposure or film processing conditions lead to lower optical densities in lesion-containing tissues [1, 2].

Consequently, attempts have been made to develop digital image receptors as a substitute for the screen-film image receptors now used in mammography [3-5]. Much of the current clinical experience with digital mammography has been derived from small field-of-view detectors that have been used for stereotactic core biopsies. These digital detectors have spatial resolutions of 5-10 lp/mm, limited primarily by the pitch of the charge-coupled device (CCD) array and the use of minimizing optics to allow a 2.5 x 2.5 cm array to cover an adequate field of view for stereotactic imaging (typically 5 x 5 cm). The issues relating to the performance of these devices have been addressed in a number of studies [6-11].

Recently, prototype whole breast digital imaging systems have been introduced for clinical testing to compare full-field digital mammography (FFDM) to screen-film mammography

(SFM). The varied designs of these digital imaging systems have been described elsewhere [3-5]. The majority of clinical testing on these systems has been unpublished studies conducted to help FFDM manufacturers gain Food and Drug Administration (FDA) approval for their devices [10]. These studies use a diagnostic mammography cohort to provide a sufficient number of cancers for testing. Unfortunately, studies based on such populations are likely to suffer from entry bias if entrance to the cohort is predicated on an abnormal screening mammogram. The study reported here eliminates that form of entry bias by enrolling a screening cohort. To our knowledge, this study is the first to compare FFDM to SFM in a screening cohort. The goal of our study is to test whether FFDM is superior to, inferior to or equivalent to SFM for screening. The study protocol is designed to minimize bias from preferential verification of findings detected on SFM ("verification bias") by recommending imaging work-up or biopsy based on positive screening results on either FFDM or SFM.

Materials and Methods

All women presenting for screening mammography at either of two university medical centers were eligible for enrollment if they were at least 40 years old, did not have breast implants, and each breast would fit entirely on a large (24 cm x 30 cm) screen-film image receptor. Each woman signed an informed consent form approved by the Institutional Review Board (IRB) of the site and the IRB of the U.S. Army Medical Research and Materiel Command.

Image Acquisition

In 4,523 cases (91%), the subject underwent FFDM immediately after her screen-film mammogram. In these cases the mammograms were acquired by the same technologist. In 422 cases (9 %), the patient presented to an outlying screening center for her SFM and underwent FFDM during a separate visit to the radiology department within three days of SFM. In these cases, a different technologist acquired the full-field digital mammogram than acquired the screen-film mammogram. Technique factors were recorded at the time of the SFM exam.

SFM was acquired on a General Electric DMR unit (GE Medical Systems, Milwaukee WI) using the Kodak Min-R 2000 screen-film system (Eastman Kodak Co, Rochester, NY). Technique factors (kVp, mAs, target and filter) for SFM were automatically selected using the Automated Optimization Parameters (AOP) feature of the unit. Contrast Mode was used when compressed breast thickness was less than 5 cm and Dose Mode was used when compressed breast thickness was greater than 5 cm. SFM quality control procedures were in accordance with the guidelines of the Mammography Quality Standards Act (MQSA).

Full-field digital mammography was performed on a prototype system utilizing an amorphous silicon area detector bonded to a CsI crystal (GE Medical Systems, Milwaukee WI). The pitch of the detector elements was 100 microns, yielding 5 line pair per millimeter limiting spatial resolution. The active area of the detector was 18 cm by 23 cm, yielding an image size of 1800 x 2304 pixels. The X-ray tube, support, and generator were identical to those in the commercial GE DMR screen-film mammography system. Each

digital detector system underwent extensive acceptance testing [8]. Ongoing quality control for the FFDM systems included daily phantom imaging and weekly flat-field detector calibration at each site.

Technique factors and breast doses for FFDM were matched to those of SFM by using the same target, filter, and kVp as used for SFM. mAs values were matched as closely as possible between the two techniques, but only discrete mAs stations were available on FFDM. In cases where the SFM mAs fell between two allowed choices on the FFDM unit, the lower mAs step was used on FFDM. Both FFDM and SFM units allowed the selection of rhodium/rhodium and molybdenum/rhodium target/filter choices in addition to molybdenum/molybdenum. Identical reciprocating grids were used for both SFM and FFDM. For the study, the average kVp was 26.90 on each modality. The average mAs was 128.6 for FFDM and 128.9 for SFM.

Rather than trying to match compression force and compressed breast thickness, the technologist was instructed to obtain the best possible positioning and compression on each modality. Compression force and compressed breast thickness were recorded for each view.

Each examination consisted of the two standard screening views of each breast: craniocaudal (CC) and mediolateral oblique (MLO) views. Both 18 x 24 cm and 24 x 30 cm screen-film image receptors were used for SFM. The FFDM prototype unit image receptor size was 18 x 23 cm. If a patient's breast was too large to fit on the 18 x 24 cm

screen-film image receptor, she was advised after her SFM that additional exposures would be needed for her FFDM to obtain coverage of each breast in each of the two standard projections. If she chose to continue in the study, her breast was imaged with FFDM by overlapping as many 18 x 23 cm views as needed to cover each breast. This typically resulted in 1 or 2 extra exposures per breast.

Image Interpretation

Screen-film images were interpreted with the routine clinical images of that day in a darkened room on a standard mammography alternator with a luminance of at least 3000 cd/m^2. The reader knew at the time of interpretation that the exam was part of the study. Digital images were interpreted in softcopy on a prototype Unix-based workstation consisting of an UltraSparc computer (Sun Microsystems, Palo Alto CA) with dual 21-inch high luminance monitors, each capable of displaying 1800x2300 pixels (Megascan Inc., North Billerica MA). These were driven by 4 megabyte video cards with 8-bit output (Dome Inc., Waltham MA). The images were interpreted without post-processing other than an initial window and level setting derived from the image histogram. A gamma function was used to map the 14-bit image data to the 8-bit display data. The radiologist had full freedom to adjust window and level and to magnify each image interactively. Magnification of each view to 2x power was typically performed for each reading. For the first approximately 200-400 cases, magnification was accomplished with a moving square showing a portion of the image magnified ("mag glass"). Subsequently, each quadrant of each image was magnified in its entirety and examined. The digital workstation was located in a darkened room away from film alternators. Comparison studies on film were viewed

on a standard lightbox placed next to the digital workstation. The lightbox was turned off during the detailed evaluation of digital images to avoid glare.

The screen-film mammogram and the digital mammogram were interpreted at the institution at which they were acquired. Interpretation was done independently by MQSA-qualified, board-certified radiologists who also interpreted clinical mammograms at that institution. For a given patient, one radiologist interpreted the SFM and another the FFDM. Each reader was blinded to the other's results and to the images from the other modality. Comparison films, prior reports and patient history were available to each radiologist for both the SFM and FFDM interpretations. At one institution, residents participated in some of the interpretations in the study, but never in both the SFM and FFDM readings of a given patient; at the other institution residents did not participate. Resident participation was not recorded, but was more likely to occur with the SFM reading than the FFDM reading. The attending radiologist, of course, gave the final interpretation. A given attending radiologist was required to interpret approximately equal numbers of SFM and FFDM examinations. Both the SFM and the FFDM interpretations were used for clinical management of the patient.

For each finding, the reader was required to give the following information:

1. A description of the type and location of the finding using BIRADS™ nomenclature. [12]

2. A BIRADS™ assessment of 0,2,3,4 or 5 (need additional imaging evaluation, benign, probably benign, suspicious abnormality or highly suggestive of malignancy). The BIRADS™ assessment category 1 (negative) was not used to describe a finding since it is reserved for cases with no findings.
3. A BIRADS™ recommendation.
4. An integer percentage probability from 0 to 100% that the finding represented cancer. This value is used for free-response ROC analysis.

Additional Imaging Evaluation of Findings

Additional imaging evaluation of a given finding was performed to establish truth for that finding in a manner that minimized bias toward either modality. Both additional mammographic views and ultrasound could be performed as indicated. Concordant findings (i.e., those detected on both modalities) needing additional mammographic views were imaged using SFM. Discordant findings (i.e., those detected on one modality only) needing additional mammographic views were imaged on the modality on which they were detected with the exception that magnification views were performed using SFM only. This exception was due to the inability to easily remove the grid on the FFDM prototype. Thus all calcifications were worked up with SFM.

Review of Discordant Cases

For all findings recalled for additional evaluation on only one of the two modalities, the two radiologists would evaluate the FFDM and SFM images side-by-side. At this time the

radiologists had the option of dismissing a finding (i.e., not working it up) if, after viewing both images, they could determine a benign cause for the finding or felt that there was no reasonable chance that it represented cancer. Dismissed findings were still counted as positives in calculations of recall rate, sensitivity and other performance measures.

Data Collection

Data were collected on paper forms and then entered into a customized database program (Microsoft Access 2.0, Belleview WA). Each patient filled out a history form detailing their risk factors and demographics. The forms were reviewed with the patient by the technologist. The technologists filled out forms detailing the reason for the exam and the technique factors. The radiologists filled out the interpretation forms. Computer entry was performed by a professional research assistant at each site.

Long-term Surveillance

Each woman in the study is followed for 1 year following her participation in the study to assess for possible development of cancers that were not detected on either modality or were detected at screening but not biopsied. Surveillance comes for most subjects when they return to the participating institution for subsequent screening mammography. Those who do not return are contacted by telephone or mail to determine if they have had subsequent mammography and whether a diagnosis of breast cancer has been made either through mammographic or clinical detection. Additionally, for subjects from one site, breast cancer cases recorded by the state tumor registry will be periodically cross-matched

against study subjects. Permission to obtain this information was included in the subject consent.

Data Analysis

For calculating sensitivity, a finding was considered to have been called on a given modality if it was assigned a recommendation for immediate work-up, such as additional imaging, obtaining prior studies for comparison or biopsy. These recommendations generally corresponded to a BIRADS™ assessment of 0, 4 or 5. A finding given a recommendation of short-interval (6 month) or routine follow-up was considered negative. These recommendations generally corresponded to a BIRADS assessment of 1, 2 or 3. Each finding was assessed separately to determine truth by imaging work-up and, if warranted, by biopsy. The final assessment of truth for a finding determined benign by additional imaging was modified if a biopsy performed on the finding within two years demonstrated cancer.

For purposes of establishing recall rate, an exam was considered positive for recall if recommended for immediate work-up or if recommended for obtaining comparison studies and those comparison studies led to additional work-up.

For calculating the number of biopsies and the positive biopsy rate, fine-needle aspiration (FNA) of solid masses, core-needle biopsy and surgical biopsy were all included. The pathologic classifications of *benign, high-risk* and *malignant* follow BIRADS™, 2nd edition

[12]. In this classification, atypical ductal hyperplasia (ADH) and lobular carcinoma in situ (LCIS) are considered high risk lesions.

Receiver operating characteristic (ROC) curves were constructed using the alternative free-response ROC (AFROC) method [13]. To utilize this approach, the value of the highest rated benign finding in each breast was assigned as the false positive level for that breast. If no finding was called in a given breast, that breast was assigned a false positive level of 0. Each cancer was assigned the rating given to it; if it was not detected it was assigned a rating of 0. Using AFROC analysis, the area under the ROC curve, A_1 is analogous to the area under a standard ROC curve, A_z, allowing the use of standard ROC analysis techniques [13-16]. The 101 point scale was used without binning. Curve areas were integrated using the trapezoidal rule.

Tests for the significance of sensitivity, recall rate, positive biopsy rate and positive predictive value were performed using McNemar's chi-square test and, if significant by that test, a Generalized Estimating Equations model implemented with PROC GENMOD in SAS (SAS Institute, Inc, Cary NC). The latter is a stricter test of significance because it takes into account the variation among readers and adds this random effect to the variance used in testing.

RESULTS

Results are given for subjects enrolled at one site as of 5/21/99 or at the other site as of 3/22/99. After these dates new workstations with higher screen resolution and automatic

post-processing were installed at each site, changing the reading environment for FFDM. In the patient cohort being reported, 4945 exams were conducted on 3890 women. For the 1055 women in the study who enrolled twice, a minimum of 11 months separated the two exams, with an average separation of 12.8 months. 2882 of the exams were conducted at one site and 2063 at the other. Results are given for the combined population except where noted.

Table 1 summarizes demographics and risk factors for the women enrolled in the study. The numbers approximate those of the screening population at each institution, with the exception of a relatively smaller number of first time screeners enrolling in the study. The biggest difference between the populations at the two institutions was that almost twice as many women at Site 1 were on hormone replacement therapy.

Table 2 compares the average compression force and compressed breast thickness for views taken on each modality. The differences are within the measurement error of the machines.

Table 3 summarizes the distribution of the radiologist's subjective rating of breast composition or mammographic density using the BIRADS™ nomenclature for each modality. There is very close agreement in this parameter between modalities.

There were 1448 findings among the study cohort for which immediate evaluation was recommended by at least one of the two readers. Of these, 507 were called positive only on FFDM, 746 were called positive only on SFM, and 195 were called positive on both.

Table 4 gives the outcome for findings by the modality on which they were called. 67 (13%) of FFDM-only findings and 14 (2%) of SFM-only findings were dismissed at discrepancy conference. All of the dismissed findings have been followed for at least one year. Three of the FFDM-only findings that had been dismissed became palpable within 1 year. Two were invasive lobular carcinomas. The third was a fibroadenoma. No other dismissed finding has undergone suspicious change or warranted biopsy for any other reason. Most were not present on the follow-up exam. Four other patients refused additional imaging and have unresolved findings. 11.1% of SFM-only findings and 7.5% of FFDM-only findings were biopsied. 15.9% of findings called on both modalities were biopsied.

Table 5 gives the distribution of finding type by modality on which it was detected. The vast majority (133/195 = 68.2%) of findings that were called on both modalities (concordant findings) were masses. Still, considering the row totals, 487 (78.5%) of the 620 mass findings were discordant, i.e., called only on one modality. Calcifications were also very likely to be called on only one modality, usually SFM. Of the 296 calcification findings in the study, 181 (61.1%) were called on SFM only, 83 (28.0%) were called only on FFDM only, and only 32 (10.8%) were called on both modalities. The greater

proportions of discordant masses, calcifications and total findings called on SFM are all statistically significant ($p<.001$).

Recall Rates

Recall rates were determined based on positive exams, rather than positive findings, consistent with the standard practice for the mammography audit [19]. A total of 568 exams were positive on FFDM for a recall rate of 11.5%. A total of 685 exams were positive on SFM for a recall rate of 13.9%. 216 of these exams were called positive on both modalities, giving a total of 1037 exams called positive for a 21.0% recall rate for study participants. The recall rates at Site 1 were 8.6% for FFDM and 11.9% for SFM. The recall rates at Site 2 were 15.5% for FFDM and 16.5 % for SFM. Findings dismissed at discrepancy conference are counted as recalls in these calculations.

Biopsied Findings and Cancers Detected

152 findings were biopsied in the study, including 23 found benign by FNA. Thirty-five cancers were diagnosed. Of these, twenty-six were invasive and 9 were DCIS. Nine of the cancers were detected on FFDM only, 10 on SFM only, and 12 on both modalities. None of these cancers was clinically suspected at the time of mammography. Four interval cancers were not detected on either modality and presented clinically within the next 11 months. Two of these were invasive lobular carcinoma and 2 were invasive ductal carcinoma. **Table 6** gives the outcome of the biopsies by modality on which the finding was detected at screening.

Table 7 gives the number of mammographically detected cancers presenting as each mammographic lesion type by the modality or modalities on which it was detected. No trend is demonstrated except for architectural distortion. Five of the 8 cancers presenting as architectural distortion were called on SFM only, versus only 2 on FFDM only. This difference is not statistically significant ($p>0.2$).

Long Term Surveillance

To date we have 1 year follow-up on 2929 of the 4945 examinations in the study. In 2861 cases the surveillance came when the subject returned to the institution for another mammogram, either as part of the continuation of this study or outside of the study. Six patients presented with clinical findings. 63 subjects have been contacted by telephone.

Sensitivity, Positive Predictive Value (PPV) and ROC Analysis

The sensitivity of FFDM in the detection of screening cancers was (21/35) 60.0%. The sensitivity of SFM was (22/35) 62.9%. As not all patients have been followed for a year to obtain complete ascertainment of interval cancers, this represents an upper bound of the sensitivity for each modality. The relative sensitivity of FFDM to SFM was (21/22) 95%.

$PPV_{screening}$ is defined as the fraction of recalled exams that led to a diagnosis of breast cancer. The $PPV_{screening}$ for FFDM was (21/568) 3.7%. The $PPV_{screening}$ for SFM was (22/685) 3.2%. The positive biopsy rate is defined as the fraction of biopsies that yielded cancer. The positive biopsy rate for all findings detected on FFDM was (21/69) 30.4%; the positive biopsy rate for all findings detected on SFM was (22/114) 19.3%.

ROC curves for the two modalities are presented in **Figure 1**. AFROC analysis was performed considering each breast separately for a total of 9716 data points (194 subjects were post-mastectomy). The area under each curve is 0.76.

Statistical Analysis

The lower recall rate for FFDM compared to SFM was statistically significant by both the McNemar's Chi-square model ($p<.0001$) and the Generalized Estimating Equations model ($p<.026$). The difference in sensitivity between the two modalities lacked statistical significance (chi-square $p > 0.5$). The higher $PPV_{screening}$ of FFDM compared to SFM was found to lack statistical significance (chi-square $p > 0.3$), as was the higher positive biopsy rate for findings detected on FFDM as compared to those detected on SFM (chi-square $p > 0.08$).

DISCUSSION

The introduction of full-field digital mammography has been eagerly awaited as a way to increase the number of cancers that can be detected by mammography. Because many cancers that cannot be detected by mammography are in dense tissue, it has been presumed that the greater contrast resolution of FFDM would enable the demonstration of at least some of these cancers. This presumption of the superiority of FFDM has been tempered by the knowledge that SFM has an advantage in high-contrast spatial resolution and may have other advantages not so easily defined, including familiarity to the radiologist. To

determine how the advantages and disadvantages of the two modalities contribute to their performance in cancer detection, we are conducting a clinical study. Because it is possible that FFDM may be either superior to, equal to or inferior to SFM, we have designed our study to be able to test for all of these possibilities. To do so necessitates that we use what is essentially a screening population, because to base enrollment on a positive clinical (screen-film) mammogram would bias the results towards SFM in terms of sensitivity and towards FFDM in terms of specificity [11]. For example had we selected only the subjects who had positive SFM exams to receive FFDM, we would have detected only 22 cancers and concluded that SFM was more sensitive than FFDM since it detected all of them whereas FFDM had only detected 12.

Unfortunately, the use of a screening population results in a relatively low number of cancers, typically 2-10 per 1000 women screened, depending on the proportion of first-time screens [19]. This low cancer rate decreases the power to detect a difference between the modalities and necessitates studying a very large population. Our cancer detection rate per thousand women screened was 4.5 for SFM. This value is in the expected range for our cohort, which consisted primarily of repeat screens [19-25]. What could not be predicted was the 39% increase in the cancer detection rate, to 6.3 per thousand, when FFDM was used in addition to SFM. This boost in cancer detection rate from the addition of FFDM increases the power of the study above that which would be predicted based on published values for yearly breast cancer incidence.

The study design maximizes statistical power by having each subject receive both SFM and

FFDM, thereby acting as her own control. This design allows the use of statistical methods for matched paired data [26,27]. The power of these methods is largely determined by the number of discordant cases. Thus the statistical power is further increased by the surprisingly large fraction of cancers, 19/31, detected on only one modality.

A potential source of bias in our study is the higher number of FFDM-only than SFM-only findings dismissed at discrepancy conference, 67 versus 14, respectively. In designing the study, we allowed a finding to not be worked up if by consensus both readers felt it could be explained as benign by looking at the other modality or was extremely unlikely to persist on additional imaging. This aspect of the protocol was included to try to reduce recalls without decreasing the rate of cancers detected. Surprisingly, out of only 67 FFDM-only findings dismissed, 2 (3.0%) turned out to be lobular carcinomas, a positivity rate higher than that in the 440 FFDM-only findings that *were* worked up. The potential for bias from this result is that there may be other undetected cancers in the remaining dismissed findings. The probability that there is even one more cancer in the group is very low, however, especially since all of these women have been followed for at least a year. Note that the dismissed findings were counted as if they had been called back. Thus the two cancers are counted as true positives for FFDM and the remaining cases as false positives.

The level of disagreement overall was surprisingly high. 821 exams had discordant interpretations, representing 17% of all exams and 79% of exams called positive on at least one modality. To the extent that this discordance is due to differences in the modality, statistical power will be increased, but to the extent it is due to factors independent of

modality, such as positioning and interpretation, statistical power will be decreased. It is well known from double-reading studies [28-30] that there is large reader variability in the interpretation of screening mammograms.

Given the advantage in contrast resolution of FFDM and the advantage in spatial resolution of SFM, one might expect that each modality would excel at detecting different types of cancers. FFDM might be expected to be better at finding densities and masses in dense tissue, while SFM might be better at detecting calcifications. While SFM did recall a larger number of calcification findings, even more than expected given its overall higher recall rate, the number of cancers presenting as microcalcifications was the same for both modalities. A higher percentage of FFDM-only calcification findings were positive at biopsy. If these trends persist, then an analysis of the reasons for the discordant readings, given by the readers at discrepancy analysis, can help to delineate whether the difference is due to FFDM being better at distinguishing benign from malignant calcifications on screening views, perhaps due to the ability to magnify the images on the workstation, or superior detection of subtle calcifications by SFM.

The only other trend in the of types of cancers detected was a larger percentage of architectural distortion cancers detected only on SFM. This trend is interesting given SFM's superior spatial resolution since detection of architectural distortion depends on resolving fine lines in breast parenchyma. This task is likely more dependent on high spatial resolution than the detection of microcalcifications, a task that requires only detecting a high-contrast focus, not on resolving its shape. More subjects are needed to

determine whether this trend in detection of architectural distortion represents a true difference between the modalities.

The only statistically significant result in our study is the lower recall rate of FFDM, due almost entirely to a lower false positive rate. Using the ROC model for evaluating a test, there are two possibilities for such a difference: either FFDM and SFM have different ROC curves, or they are being interpreted at different points along the same curve, with SFM's operating point being to the right of FFDM's. Although a difference in false positive rate without a difference in sensitivity implies that the two modalities have different ROC curves, because the power to detect a difference in sensitivity is relatively low, we cannot yet exclude a small difference in sensitivity that would explain the difference in false positive rate as a shift along the same ROC curve. To help distinguish operating point shifts from differences in ROC curves, ROC data were collected on a 101 point scale separately from the BIRADS™ assessments and recommendations. The finding of nearly identical ROC curves for FFDM and SFM supports that the difference in false positive rate reflects a shift of operating point along the same ROC curve. Such a shift could be caused by a difference in the reading conditions between SFM and FFDM, so these conditions were kept as equivalent as possible. One difference at one site was the participation of residents, which, anecdotally, was more likely to occur on SFM than on FFDM. The other site had no resident participation. Since resident participation was not tracked, whether the participation of residents had a noticeable effect on recall rates cannot be determined. Both sites had lower recall rates on FFDM, but the effect was larger at the site that allowed resident participation. Because of this and other possible unappreciated differences in the

reading environments, whether the lower recall rate on FFDM is a property of the modality that will be observed in clinical practice or is an artifact of the experimental situation cannot be definitely determined from our data.

CONCLUSION

A prospective study comparing FFDM to SFM in a screening cohort has been conducted at two mammography sites. Results indicate that the two modalities have indistinguishable sensitivities for cancer detection and indistinguishable ROC curves but that FFDM has a statistically significantly lower recall rate than SFM. Statistical power is limited by the inherently small number of cancers in a screening population. For this reason, a second phase of this study, currently underway, includes a third site and will accrue more subjects to increase statistical power. Additionally, a larger, multi-institutional trial with a similar design is being planned by the American College of Radiology Investigative Network (ACRIN).

Acknowledgments:

We would like to acknowledge the important contributions of Virginia Vance and Vanessa Brown as study coordinators. We are also grateful to the radiologic technologists who have pioneered digital mammography at our two institutions: Melody Lisne, Mary Heck, Cindy Hauger, Mary Benedict, Cheryl Porcaro, Laurie Theard and Candace Kennedy.

REFERENCES

1. Young KC, Wallis MG, Ramsdale ML. Mammographic film density and detection of small breast cancers. Clinical Radiology 1994; 49: 461-465.

2. Robson KJ, Kotre CJ, Faulkner K. The use of a contrast-detail tests object in the optimization of optical density in mammography. The British Journal of Radiology 1995; 68: 277-282.

3. Yaffe MJ, Nishikawa RM, Maidment ADA, Fenster A, "Development of a digital mammography system". SPIE Proc. 1988; 914:182-188.

4. Yaffe MJ, Rowlands JA. X-ray detectors for digital radiography. Phys Med Biol 1997 Jan;42(1):1-39.

5. Yaffe MJ. Rowlands JA. X-ray detectors for digital mammography. Phys. Med. Biol. 1997; 42: 1-39.

6. Karellas A, Hendrick RE. Equipment: digital mammography. In Breast Imaging – Categorical Course Syllabus. eds., PJ Dempsey and B Monsees, American Roentgen Ray Press, 1999: 1-9.

7. Kimme-Smith C, Beifuss M, Williams M, Bassett LW. Establishing minimum performance standards, calibration intervals, and optimal exposure values for a whole breast digital mammography unit. Medical Physics 25(12): 2410-2416, 1998.

8. Vedantham S, Karellas A, Suryanarayanan S, et.al. Full-breast digital mammographic imaging with an amorphous silicon-based flat-panel detector: Physical characteristics of a clinical prototype. Medical Physics 27(3): 558-566, 2000.

9. Hendrick RE, Berns E, Chorbajian B, Choi K, Lewin JM, Sisney GA, et.al. Low-contrast lesion detection: comparison of screen-film and full-field digital mammography. (abstract) Radiology 205(P): 274, 1997B.

10. Pisano E, Yaffe MJ, Hemminger B, Hendrick RE, et.al. Current status of full-field digital mammography. Submitted to Academic Radiology, 1999.

11. Lewin JM, Hendrick RE, D'Orsi CJ, Moss LJ, Sisney GA, Karellas A. Clinical evaluation of a full-field digital mammography prototype for cancer detection in a screening setting – work in progress. Radiology 209(P): 238, 1998.

12. American College of Radiology (ACR). Breast imaging reporting and data system (BI-RADS™). Second Edition. Reston [VA]: American College of Radiology; 1995.

13. Chakraborty DP, Loek HLW. Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology 1990; 174: 873-881.

14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29-36.

15. Chakraborty DP. Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Med. Phys. 1989; 16: 561-568.

16. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839-843.

17. Dorfman DD, Alf E. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals: rating-method data. J. Math. Psychol. 1969; 6: 487-496.

18. Obuchowski NA, McClish DK. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. Statistics in Medicine 1997; 16: 1529-

1542.

19. Bassett LW, Hendrick RE, Bassford TL, et al. *Quality Determinants of Mammography. Clinical Practice Guideline No. 13.* AHCPR Publication No. 95-0632. Rockville MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Dept of Health and Human Services. Oct 1994.

20. Bird RE. Low-cost screening mammography: report on finances and review of 21,716 consecutive cases. Radiology 1989; 171:87-90.

21. Braman DM, Williams HD. ACR accredited suburban mammography center; three year results. J Fla Med Assoc 1989; 76:1031-4.

22. Linver MV, Paster S, Rosenberg RD, Key CR, Stidley CA, King WV. Improvement in mammography interpetation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases. Radiology 1992; 184:39-43.

23. Lynde JL. A community program of low-cost screening mammography: the results of 21,141 consecutive examinations. South Med J 1993; 86:338-343.

24. Robertson CL. A private breast imaging practice: medical audit of 25,788 screening and 1077 diagnostic examinations. Radiology 1993; 187:75-79.

25. Sickles EA. Quality assurance: how to audit your own mammography practice. Radiol Clin North Am 1992; 30:265-275.

26. Siegel S. Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

27. Liang KY, Zeger SL. Regression analysis for correlated data. Annual Rev Public Health 1993; 14:43-68.

28. Anderson EDC, Muir BB, Walsh JS, Kirkpatrick AE. The efficacy of double reading mammograms in breast screening. Clin Radiol 1994; 49:248-251.

29. Thurfjell EL, Lerneval KA, Taube AAS. Benefit of independent double reading in a population-based mammography screening program. Radiology 1994; 191:241-244.

30. Beam CA, Sullivan DC, Layde PM. Effect of human variability on independent double reading in screening mammography. Acad Radiol 1996; 3:891-897.

TABLE 1
Demographics And Risk Factors For The Study Cohort By Site And Combined

	Site 1	Site 2	Combined
Number of exams	2882	2063	4945
Mean age +/- standard dev'n	56.3 +/- 10.0	54.6 +/- 9.4	55.5 +/- 9.8
First mammogram	40 (1.4%)	7 (0.3%)	47 (0.9%)
On hormone replacement	1702 (59.0%)	716 (34.7%)	2418 (48.9%)
Personal history of breast cancer	159 (5.5%)	197 (9.5%)	356 (7.2%)
Strong family history[1]	206 (7.2%)	114 (5.5%)	320 (6.5%)
Intermediate family history[2]	335 (11.6%)	302 (14.6%)	637 (12.9%)
Weak family history[3]	649 (22.5%)	358 (17.3%)	1007 (20.4%)
Nulliparous	634 (22.0%)	327 (15.8%)	965 (19.4%)
Late child bearing	362 (12.5%)	167 (8.1%)	529 (10.7%)

[1]Mother, sister, daughter, premenopause or postmenopausal multiple first-degree relatives
[2]Mother, sister, daughter postmenopause
[3]Aunt, grandmother, cousin

TABLE 2
Compression Force And Compressed Breast Thickness By Modality
(average ± standard deviation)

	FFDM	SFM
Compression force (decaNewtons)	11.2 ± 4.2	11.1 ± 4.1
Compressed breast thickness (mm)	51.0 ± 14.9	49.9 ± 13.6

TABLE 3
Radiologist Rating Of Breast Composition By Modality

	FFDM	SFM
Fatty	465 (9.5%)	449 (9.2%)
Scattered densities	2226 (45.6%)	2250 (46.1%)
Heterogeneously dense	1726 (35.3%)	1725 (35.3%)
Extremely dense	466 (9.6%)	449 (9.4%)

Based on 4883 cases in which ratings were available for both modalities.

TABLE 4
Outcome of Findings by Modality

Outcome	findings detected on FFDM only	findings detected on SFM only	findings detected on both
Dismissed at Discrepancy Conference	64 *	14	-
Unresolved	0	3	1
Benign or Probably Benign after Imaging Work-up	405	646	163
Biopsied	38	83	31
TOTAL	507	746	195

* Does not include 3 discarded findings that later became palpable and were biopsied

TABLE 5
Distribution of Finding Type by Modality

Finding Type	findings detected on FFDM only (%)	findings detected on SFM only (%)	findings detected on both (%)	**Total (%)**
Mass	203 (32.7)	284 (45.8)	133 (21.5)	**620 (100)**
Asymmetric Density	166 (41.1)	214 (53.0)	23 (5.9)	**406 (100)**
Calcification	83 (28.0)	181 (61.1)	32 (10.8)	**296 (100)**
Architectural distortion	42 (43.7)	50 (52.1)	4 (4.2)	**96 (100)**
Other / not specified	12 (40.0)	16 (53.3)	2 (6.7)	**30 (100)**
Total	**507 (35.0)**	**746 (51.5)**	**195 (13.5)**	**1448 (100)**

TABLE 6
Biopsy Results

	findings detected on FFDM only	Findings detected on SFM only	findings detected on both
Benign	24	72	19
High risk	5 [1]	1 [2]	0
Malignant (invasive)	9 (7)	10 (7)	12 (8)

[1] 3 cases of ADH; 2 of LCIS
[2] LCIS

TABLE 7
Mammographically Detected Cancers By Mammographic Lesion Type

Lesion Type	Number detected on FFDM - only	Number detected on SFM - only	Number detected on both modalities
Mass	2	2	4
Asymmetric Density	2	1	1
Calcification	3	2	6
Architectural Distortion	2	5	1
Total	**9**	**10**	**12**

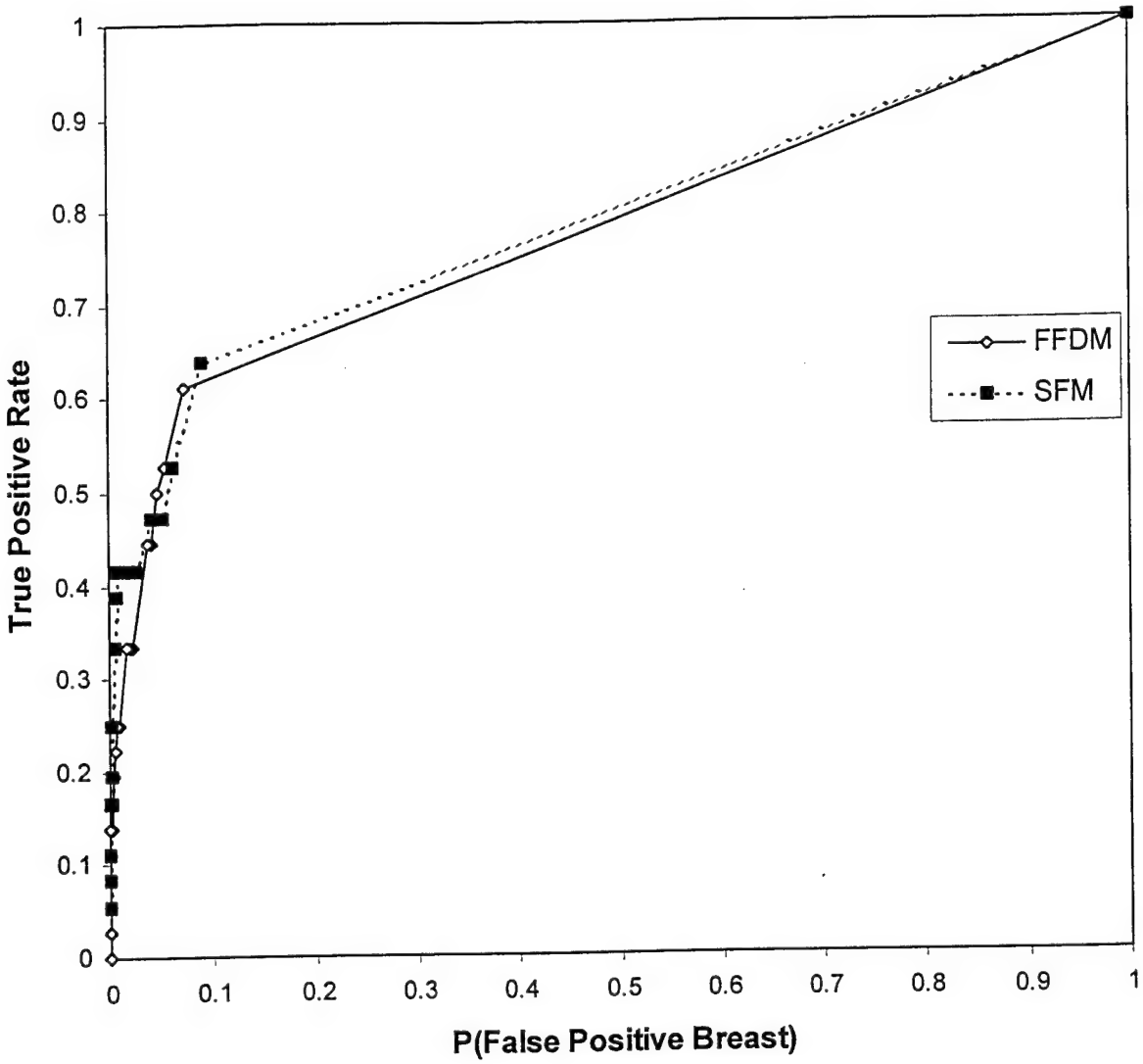

Figure 1. Alternative free-response ROC (AFROC) curves for SFM and FFDM plotted using a rating scale from 0-100. The x-axis scale is the probability of a false positive finding being called on the two screening views of a given breast. This is analogous to the false positive rate in a standard ROC experiment. The areas under the curves are 0.78 for FFDM and 0.79 for SFM. The difference in areas is not statistically significant ($p>0.4$).

ID: Date: / / rev. 12/1/98

Rad ID(s): _____ _____ _____ _____

Exam ID# : _____ Finding#: _____

UCHSC/UMMC/NNMC Full-Field Digital Mammography Study

DISCREPANCY FORM:

Relative Conspicuity: 0 1 2 3 4 5 6 7 8 9 10

Lesion only Visible on Dig — Equally Visible — Lesion only Visible on Film

DISCREPANCY REASON(S):

Primary:_____ Secondary (optional):_____

V - Visibility/Conspicuity

1.Contrast: V1u- unsure cause V1m- modality difference
V1c- compression difference V1p- fortuitous positioning

2. Sharpness: V2d - Sharpness, greater on digital V2f - Sharpness, greater on film

3. Positioning: V3s - Positioning, due to small dig detector size V3p - Positioning, other

4. Artifacts: V4sd - Artifact simulates finding on dig V4hd - Artifact hides finding on dig
V4sf - " film V4hf - " film

5. Motion: V5d - Patient motion on dig V5f - Patient motion on film

6. Other: V6 - other cause for visibility/conspicuity difference (comment on bottom of form)

7. Magnification: V7d - ability to magnify on dig workstation

A - Appearance:

A1d - mass margin/shape more suspicious on digital
A1f - margin/shape more suspicious on film
A2d - calcification forms more suspicious on digital
A2f - calc forms more suspicious on film
A3d - calcification distribution more susp. on digital
A3f - calc distribution more susp. on film
A4d - more calcs visible on digital
A4f - more calcs visible on film
A5d - greater change from comparison on digital
A5f - greater change on film

I - Interpretation

I1a - Difference of opinion, Minor I1b - Difference of opinion, Major
I2d - Error in interpretation, Digital I2f - Error in interpretation, Film
I3d - Error in detection, Digital I3f - Error in detection, Film
I4d - Dig assessed as more changed from comparison I4f - Film assessed as more changed
I5d - Did not have comparisons for dig (only) I5f - Did not have comparisons for film (only)
I6w - Poor use of window/level on dig I6c - Poor use of other workstation controls
N0a - Other reason or Not Sure as to reason (comment on back - optional)

Clinical BIRADS Assessment (0-5): ____ Clinical Recommendation*: _____

* N=routine; O=get comparisons; AV=add'l views U=ultrasound; F=6m follow up; H=Bx; D=clinical eval

Breast Imaging Using an Amorphous Silicon-Based Full-Field Digital Mammographic System: Stability of a Clinical Prototype

Srinivasan Vedantham, Andrew Karellas, Sankararaman Suryanarayanan, Carl J. D'Orsi, and R. Edward Hendrick

An amorphous silicon-based full-breast imager for digital mammography was evaluated for detector stability over a period of 1 year. This imager uses a structured CsI:Tl scintillator coupled to an amorphous silicon layer with a 100-micron pixel pitch and read out by special purpose electronics. The stability of the system was characterized using the following quantifiable metrics: conversion factor (mean number of electrons generated per incident x-ray), presampling modulation transfer function (MTF), detector linearity and sensitivity, detector signal-to-noise ratio (SNR), and American College of Radiology (ACR) accreditation phantom scores. Qualitative metrics such as flat field uniformity, geometric distortion, and Society of Motion Picture and Television Engineers (SMPTE) test pattern image quality were also used to study the stability of the system. Observations made over this 1-year period indicated that the maximum variation from the average of the measurements were less than 0.5% for conversion factor, 3% for presampling MTF over all spatial frequencies, 5% for signal response, linearity and sensitivity, 12% for SNR over seven locations for all 3 target-filter combinations, and 0% for ACR accreditation phantom scores. ACR mammographic accreditation phantom images indicated the ability to resolve 5 fibers, 4 speck groups, and 5 masses at a mean glandular dose of 1.23 mGy. The SMPTE pattern image quality test for the display monitors used for image viewing indicated ability to discern all contrast steps and ability to distinguish line-pair images at the center and corners of the image. No bleeding effects were observed in the image. Flat field uniformity for all 3 target-filter combinations displayed no artifacts such as gridlines, bad detector rows or columns, horizontal or vertical streaks, or bad pixels. Wire mesh screen images indicated uniform resolution and no geometric distortion.

KEY WORDS: breast imaging, digital mammography, physics, image quality, flat panel technology, linearity, modulation transfer function.

SCREEN-FILM TECHNOLOGY currently is the standard in full-breast imaging for the detection of nonpalpable breast cancer. Advancements in digital technology in the past decade have made it feasible to obtain high-quality images using digital detectors, and earlier this year the US Food and Drug Administration approved the first full-breast digital imaging system. Screen-film mammography is very effective for screening, but it has well-known limitations with regard to dynamic range, contrast, and lack of convenient options for postprocessing of images. Digital detectors have the theoretical capability to overcome some of these fundamental limitations. Digital detectors also show the potential to provide high detection efficiency, high dynamic range, capability of contrast enhancement[1] and postprocessing capabilities including computer-aided diagnosis.[2-7] In addition, with the advent of digital detectors it is feasible to perform novel imaging techniques like tomosynthesis,[8-11] digital subtraction,[12] and dual-energy mammography.[13,14]

Different modes of electronic detection like charge-coupled devices[15-19] (CCDs) and image intensifiers have been used in the past. A detailed study on a CCD-based whole-breast digital mammography unit with emphasis on establishing performance standards based on long-term stability also has been reported.[20] Improvements in thin-film transistor (TFT) technology and the ability to manufacture large flat panel semiconductor monolithic arrays have resulted in increased interest in utilizing these arrays for full-breast digital mammography. Currently, amorphous silicon[21-27] and amorphous selenium[28,29] semiconductors are sub-

From the Department of Radiology, UMass Memorial Health Care, University of Massachusetts Medical School, Worcester, MA, and the Radiology Department, University of Colorado Health Science Center, Denver, CO. (Presently at Lynn Sage Comprehensive Breast Center, Northwestern University Medical School, Chicago, IL.)

This work was supported in part by US Army grant DAMD17-96-C-6104 to the University of Colorado Health Sciences Center, and in part by the grant R01CA59770 from the National Cancer Institute to the University of Massachusetts Medical School. The full-field digital mammography detector was developed independently by GE Corporate Research and Development with partial support from the National Cancer Institute grant 5R01CA60183. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of NCI, NIH, or US Army.

Address reprint requests to Andrew Karellas, PhD, Professor of Radiology, Department of Radiology, Room S2-836, University of Massachusetts Medical School, UMass Memorial Health Care, 55 Lake Ave North, Worcester, MA 01655.

0897-1889/00/1304-0002$10.00/0
doi:10.1053/jdim.2000.17764

jects of active research for imaging applications. Amorphous silicon-based detectors use a scintillator for conversion of x-ray photons to light photons; the generated light photons subsequently are detected by the amorphous silicon photodiode unlike the direct conversion photoconductor used by amorphous selenium-based detectors.

This study characterizes the stability of an amorphous silicon-based clinical prototype flat panel imager (GE Medical Systems, Milwaukee, WI), which underwent technical and clinical evaluation at the University of Massachusetts Medical School and the University of Colorado Health Sciences Center. Although this study characterizes the stability of a clinical feasibility system, the stability of a clinical system, which has a larger pixel matrix (1920 × 2304), is the subject of an ongoing study.

MATERIALS AND METHODS

The full-breast digital mammography imager used in this study consists of a thallium-doped CsI:Tl scintillator and an amorphous silicon photodiode array and incorporates special-purpose readout electronics. Light created from the interaction of x-ray photons in the scintillator travels down the columnar crystalline structure of the scintillator, which is in contact with a 2-dimensional array of amorphous silicon photodiodes and thin-film transistors. Light exiting from the scintillator is detected by the monolithic thin film flat panel array, which consists of a matrix of 1800 × 2304 pixels that are 100 μm in pitch. The specifications of the flat panel imager are presented in Table 1. Each pixel in the array is an individually addressable light detector. The electrical signals of all pixels are individually read out and digitized to 16-bit digital values in 300 ms by special-purpose low-noise electronics,[30] which are located inside the image receptor assembly. The schematic of the detector is shown in Fig 1. The imager is integrated into a prototype digital mammography system based on a multipulse high-frequency x-ray generator (Senographe DMR, GE Medical Systems). This system uses an x-ray tube with a selectable dual track target, either molybdenum (*Mo*) or rhodium (*Rh*) and selectable filtration of *Mo* or *Rh*. The stability of the system was quantified in terms of the following metrics: conversion factor (number of electrons detected per incident x-ray), modulation transfer function (MTF), detector signal-to-noise ratio (SNR) measurement, flat field uniformity, and ACR accreditation phantom scores.

Table 1. Specifications of the Amorphous Silicon-Based Flat Panel Full-Breast Digital Mammographic Imager

Image area	18 cm × 23 cm
Image matrix	1800 × 2304
Scintillator	CsI: Tl
Pixel pitch	100 microns

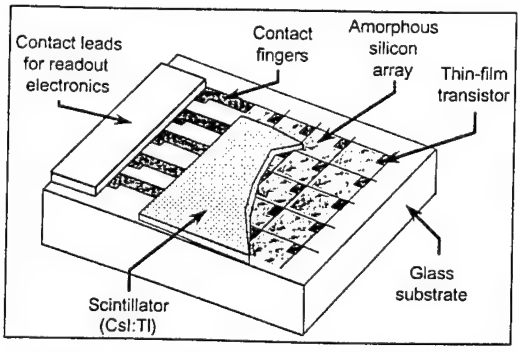

Fig 1. Schematic of the amorphous silicon-based flat panel full-breast digital mammographic imager.

Conversion Factor

The conversion factor is defined as the mean number of electrons detected per incident x-ray. The number of incident photons was estimated using an x-ray spectral model built into the system as part of the quality control tests and provided by the manufacturer. Because this x-ray spectral model addresses only the 30 kVp, *Rh* target and *Rh* filter spectrum, measurements were performed with an exposure of 30 kVp, *Rh* target and *Rh* filter at 20 mAs as per the manufacturer recommended procedure. The mean number of generated electrons was computed using the algorithm provided with the system, which takes into account the mean number of digital exposure units of the central 256 × 256 pixels of the full-breast detector and the gain of the amplifier. This measurement was performed over a period of 1 year on a monthly basis to assess variations in conversion factor.

Presampling MTF

The presampling MTF of the system was characterized using the edge response function (ERF) method. The ERF was obtained using a 0.125-mm thick tungsten bar placed at a slight angle (<5°) as shown in Fig 2. Because each pixel represents a sample of the ERF at a distance equal to the center of the pixel to the edge, the oversampled ERF was obtained by plotting the pixel intensity from the edge. This technique is similar to the methodology used by Fujita et al[31] to obtain the oversampled LSF. The line-spread function (LSF) was derived from the ERF by differentiation. The modulus of the Fourier transform of the LSF provided the presampling MTF of the system. The ERF was obtained by exposing the tungsten bar at 30 kVp, *Rh* target, *Rh* filter, and 20 mAs as per manufacturer's recommendations. This measurement was performed over a period of 1 year on a monthly basis to assess variations in MTF. An alternate technique using a 10-micron slit to obtain the LSF also was used to confirm the results obtained.[26]

Detector Linearity and Sensitivity

The linear response was measured at 28 kVp, *Mo* target, *Mo* filter, and added filtration of 4.5 cm of Lucite, by averaging the pixel intensities over a 256 × 256 area located at the center of the image at various exposure levels. The sensitivity of the

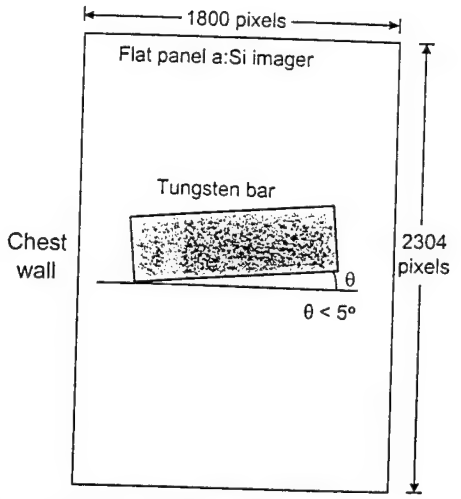

Fig 2. Illustration of the angular placement of the tungsten bar to obtain the edge response function (ERF) for modulation transfer function (MTF) measurements.

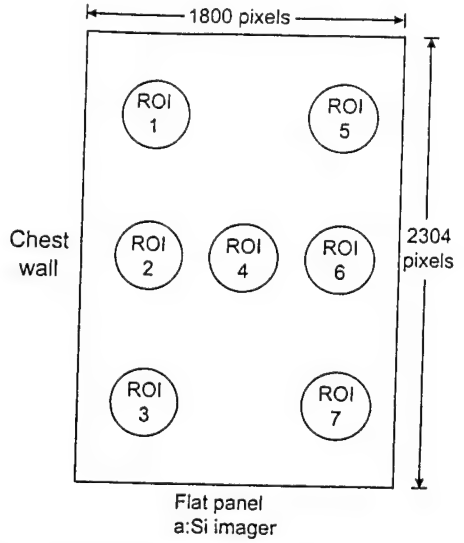

Fig 3. Illustration of the location of the ROI for SNR measurements.

system was determined by evaluating the slope of the linear response curve.[26]

Detector Signal-to-Noise Ratio

The detector SNR was measured at 7 locations across the imager for all 3 target-filter combinations. A 1-inch thick uniform sheet of Lucite covering the entire detector was placed on the breast support plate. The exposure techniques for the three target-filter combinations are shown in Table 2. Exit exposures (exposure on to the breast support plate) were measured with the 1-inch-thick Lucite sheet elevated above the ion chamber with a 10-cm air gap between the 2 objects similar to the technique used by Kimme-Smith et al.[20] The acquired images were dark-image subtracted and flat-field corrected[25] and viewed at the image-review station (Sun Microsystems, Palo Alto, CA). The image-review station is equipped with dual CRT displays driven by special purpose display-driver boards (Dome Imaging Systems, Inc, Waltham, MA). The mean pixel intensity, μ (digital values) and standard deviation, σ, were measured using a 539 mm^2 region of interest (ROI) at 7 locations of the image as shown in Fig 3. The SNR was obtained by dividing the mean pixel intensity by the standard deviation (μ/σ).

Table 2. Exposure Techniques Used for All 3 Target-Filter Combinations for SNR Measurements and Flat-Field Uniformity Measurements

Target-Filter	Tube Voltage (kVp)	Tube Current × Time (mAs)	Exit Exposure (mR)
Mo-Mo	25	110	63.9
Mo-Rh	28	80	72.0
Rh-Rh	32	40	67.1

NOTE. The exit exposure represents the exposure on to the breast support plate after transmitting through 1-inch of Lucite.

ACR Accreditation Phantom Scores

An American College of Radiology (ACR) Mammography accreditation phantom (Model: 18-220, Nuclear Associates, Cleveland, OH) was imaged using the following technique factors: *Mo/Mo,* 25 kVp, 100 mAs, large focal spot, 660 mm source-to-image distance (SID), with antiscatter grid. This technique corresponds to a mean glandular dose of 1.23 mGy. The images were viewed at the image-review station. Optimum windowing and leveling, and magnification of the image up to a factor of 1.7 was utilized for scoring the phantom. Phantom scoring was performed according to ACR guidelines.[32] Then the images were printed onto a dry laser compatible film (DryView Laser Imaging Film, Eastman Kodak Co, Rochester, NY) using a dry-film laser printer (DryView 8600, Eastman Kodak Co) and scored according to ACR guidelines.[32]

Qualitative Measures

In addition to these quantitative measurements, qualitative measurements to characterize the stability of the system to provide consistent image quality also were performed. These studies included flat-field uniformity, geometric distortion, and Society of Motion Picture and Television Engineers, White Plains, NY (SMPTE) test pattern image quality for the display monitors used for patient image viewing.

The flat-field uniformity of the detector was studied by imaging a 1-inch thick uniform sheet of Lucite covering the entire detector placed on the breast support plate, using the technique factors listed in Table 2 for all 3 target-filter combinations. The images were studied at the image-review station for artifacts such as gridlines, bad detector rows or columns, horizontal or vertical streaks, and bad pixels.

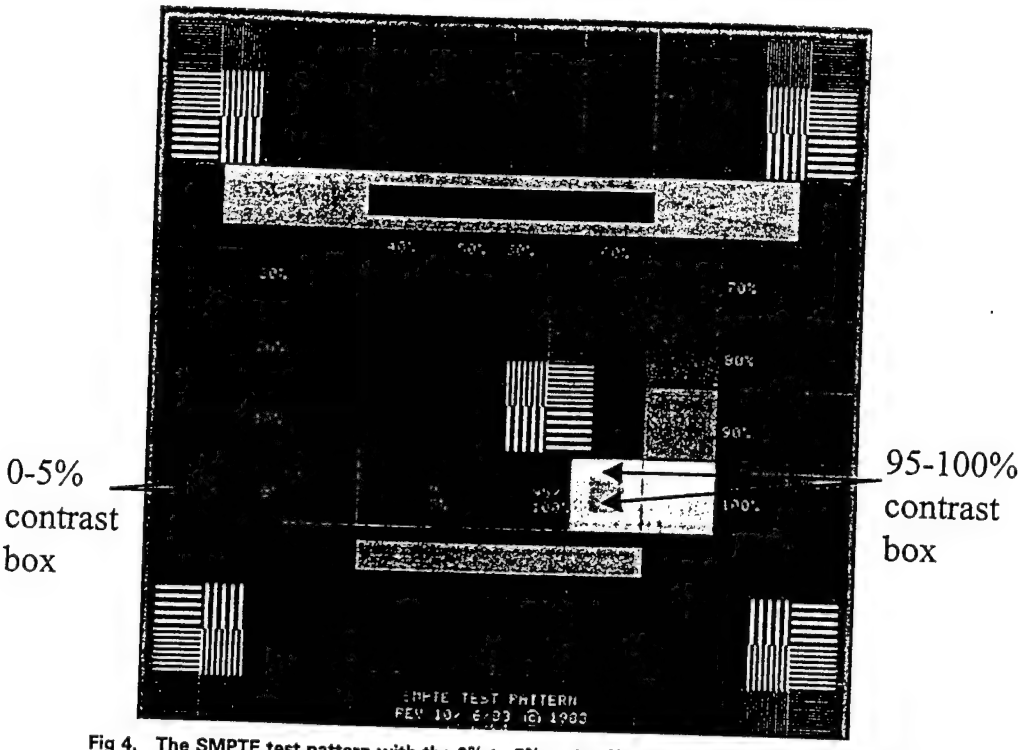

Fig 4. The SMPTE test pattern with the 0% to 5% and 95% to 100% contrast boxes labeled.

Geometric distortion was evaluated by imaging a wire mesh screen at 25 kVp, *Mo/Mo,* 50 mAs, 660 mm SID, and with the antiscatter grid. The image was visualized at the image-review station and evaluated for resolution uniformity and distortion.

The SMPTE pattern image quality was used to assess the stability of the CRT displays used for patient image viewing at the image-review station. The photometer test tool provided by Dome was utilized for generating the SMPTE pattern. An image of the SMPTE test pattern is shown in Fig 4 with the 0% to 5% and 95% to 100% contrast boxes labeled. An extensive description of the use of SMPTE test pattern images for evaluating soft copy devices was given by Nawfel et al.[33] The SMPTE pattern image was studied for the following features: ability to discern 0% to 5% contrast box and 95% to 100% contrast box, ability to distinguish line-pair images at the center and corners of the SMPTE pattern, and presence of any bleeding effects.

RESULTS AND DISCUSSION

The conversion factor measured over a period of 1 year as shown in Fig 5 was stable and indicated very little variability. The maximum variation from the mean was less than 0.5%. The mean presampling MTF measured is shown in Fig 6. The error bars represent the maximum variations observed over the period of 1 year at discrete spatial frequencies. The presampling MTF measured over a period of 1 year at 1, 2, 3, 4, and 5 cycles per millimeter are shown in Fig 7. The maximum variation from the mean at these discrete spatial frequencies was less than 3% over a period of one year. This variation can be attributed partially to experimental error.

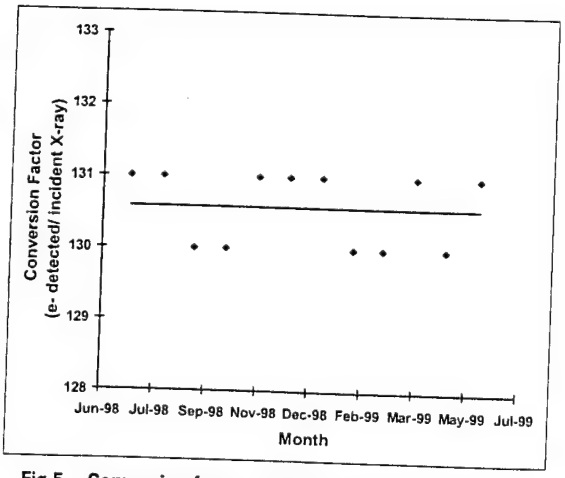

Fig 5. Conversion factor measurements over a period of 1 year performed on a monthly basis. Maximum variation was less than 0.5% during this period.

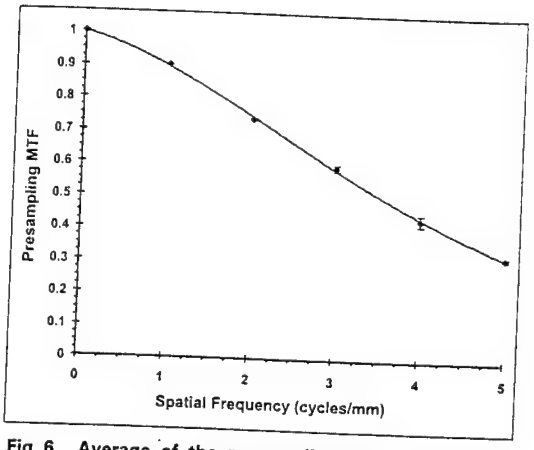

Fig 6. Average of the presampling MTF measurements made over a period of 1 year on a monthly basis. The error bars represent the maximum variations observed over this duration at discrete spatial frequencies.

The measured detector linearity is shown in Fig 8. The error bars represent the maximum deviations from the average of the mean pixel intensity (digital units) over a 256 × 256 ROI observed over a period of 1 year. From the linearity measurements the mean sensitivity of the system was found to be 16.3 digital units per milliroentgen per pixel. Maximum variation in linearity and sensitivity was less than 5% over this period.

The signal, defined as the mean of each of the 539 mm^2 ROI as shown in Fig 3, varied by less than 5% (4% for *Mo-Mo,* 4% for *Mo-Rh,* and 2% for *Rh-Rh*) over the 7 locations for all 3 target-filter combinations. The coefficient of variation defined as the ratio of the standard deviation to the mean ($\sigma/\mu \times 100\%$),[20] in SNR measurements over the 7 locations of the detector was less than 7% for the *Mo-Mo,* less than 9% for the *Mo-Rh,* and less than 4% for the *Rh-Rh* target-filter combinations. As an example, one such SNR measurement is shown in Fig 9. The SNR measurements seem to indicate a trend in which the SNR decreases from the edge of the detector proximal to the chest wall (ROI locations 1, 2, and 3) toward the edge of the detector distal to the chest wall (ROI locations 5, 6, and 7). To verify if this trend was caused by the heel effect, the signal amplitudes of the ROIs 2, 4, and 6 were analyzed for each of the 3 target-filter

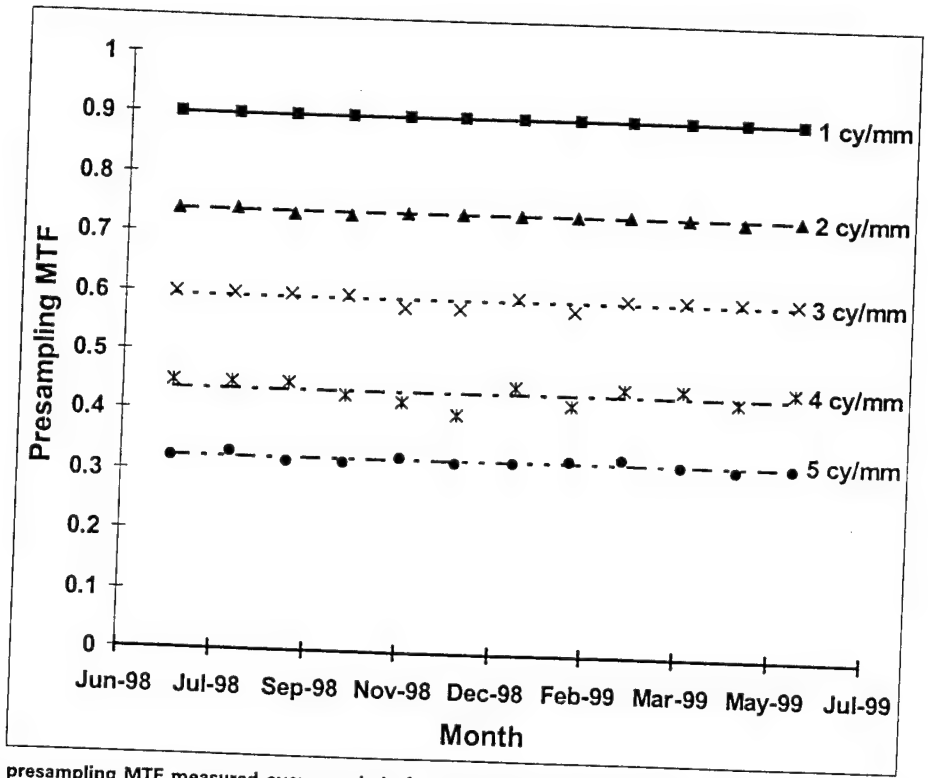

Fig 7. The presampling MTF measured over a period of 1 year at discrete spatial frequencies of 1, 2, 3, 4, and 5 cycles per millimeter (cy/mm). The maximum variation from the mean at these discrete spatial frequencies was less than 3% over this period.

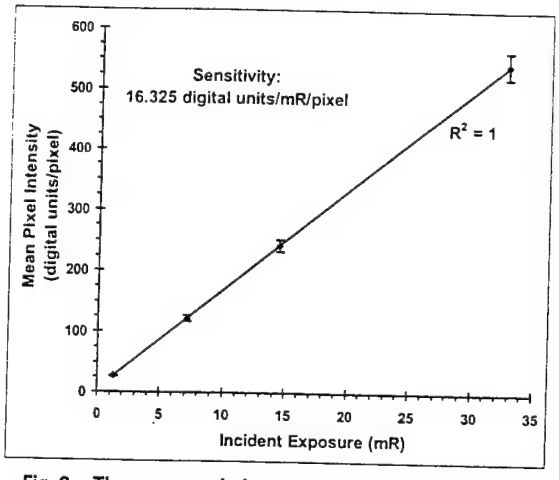

Fig 8. The measured detector linearity. The error bars represent maximum deviations from the average of the mean pixel intensity (digital units) of a 256 pixels × 256 pixels ROI observed over a period of 1 year.

combinations. The signal amplitudes after flat-field correction from these 3 ROIs did not indicate any fall-off trend. This suggests that the flat-field correction technique was adequate to suppress the heel effect. The heel effect would result in higher exposure at the edge proximal to the chest wall compared with the edge distal to the chest wall. This would result in higher signal amplitudes at the edge proximal to the chest wall compared with the edge distal to the chest wall. Flat-field correction techniques would compensate this trend resulting in an almost uniform signal throughout the entire field. However, under quantum-limited operating conditions, a difference in exposure levels between the edge proximal to the chest wall and the edge distal to the chest wall would lead to a trend in the SNR proportional to the square root of the difference in the number of x-ray photons. Hence, the ratio of the SNR obtained from ROIs 2 and 6 were compared with the square root of the ratio of the exposure levels at ROIs 2 and 6. This ratio indicated that the measured SNR fall-off trend was within 5% of the expected SNR fall-off trend. This suggests that the flat-field correction technique was adequate in flattening the field (signal) and the difference in the exposure levels between the edge proximal to the chest wall and the edge distal to the chest wall is a major contributor to the observed SNR fall-off trend. The mean SNR over the 7 locations was 121 for the *Mo-Mo*, 138 for the *Mo-Rh*, and 130 for the *Rh-Rh*. The increases (%) in mean SNR with the *Mo-Rh* and *Rh-Rh* target-filter combinations compared with the *Mo-Mo* target-filter combination were 14% and 7.5%. The improvement in the mean SNR for the *Mo-Rh* and *Rh-Rh* target-filter combinations are attributed partially to the higher exit exposures (12.6% higher for the *Mo-Rh* and 5% higher for the *Rh-Rh* target-filter combinations compared with the *Mo-Mo* target-filter combination).

Images of the ACR accreditation phantom scored both with the hard copy (printed film) and the soft copy (CRT display) showed the ability to resolve 5 fibers, 4 speck groups, and 5 masses at breast doses comparable to that of screen-film phantom images. These images did not show any variability over the 1-year period. An ACR accreditation phantom image acquired at 25 kVp, 100 mAs, *Mo-Mo* target-filter combination, 660 mm SID, with antiscatter grid corresponding to a mean glandular dose of 1.23 mGy is shown in Fig 10. The process of transferring the image on to a paper print resulted in a slight degradation, but the hard copy (printed film) and soft copy (CRT display) images indicated the capability of resolving 5 fibers, 4 speck groups, and 5 masses.

The SMPTE pattern image quality test indicated ability to discern all contrast steps, 0% to 5% contrast box, 95% to 100% contrast box, and ability to distinguish line-pair images at the center and corners of the SMPTE pattern image. No

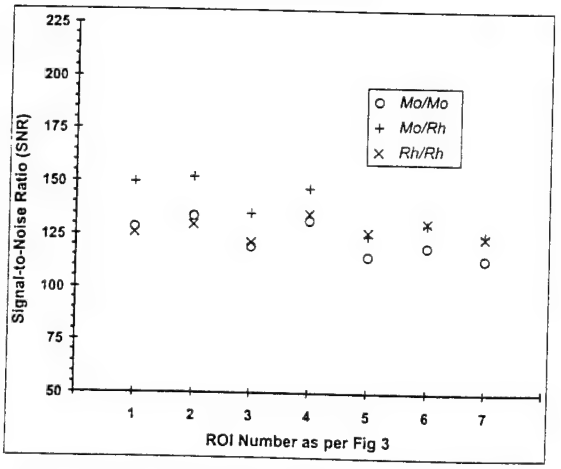

Fig 9. SNR measurements over the 7 locations of the detector as shown in Fig 3 for all 3 target-filter combinations. Observed variation was less than 12% from the central ROI 4.

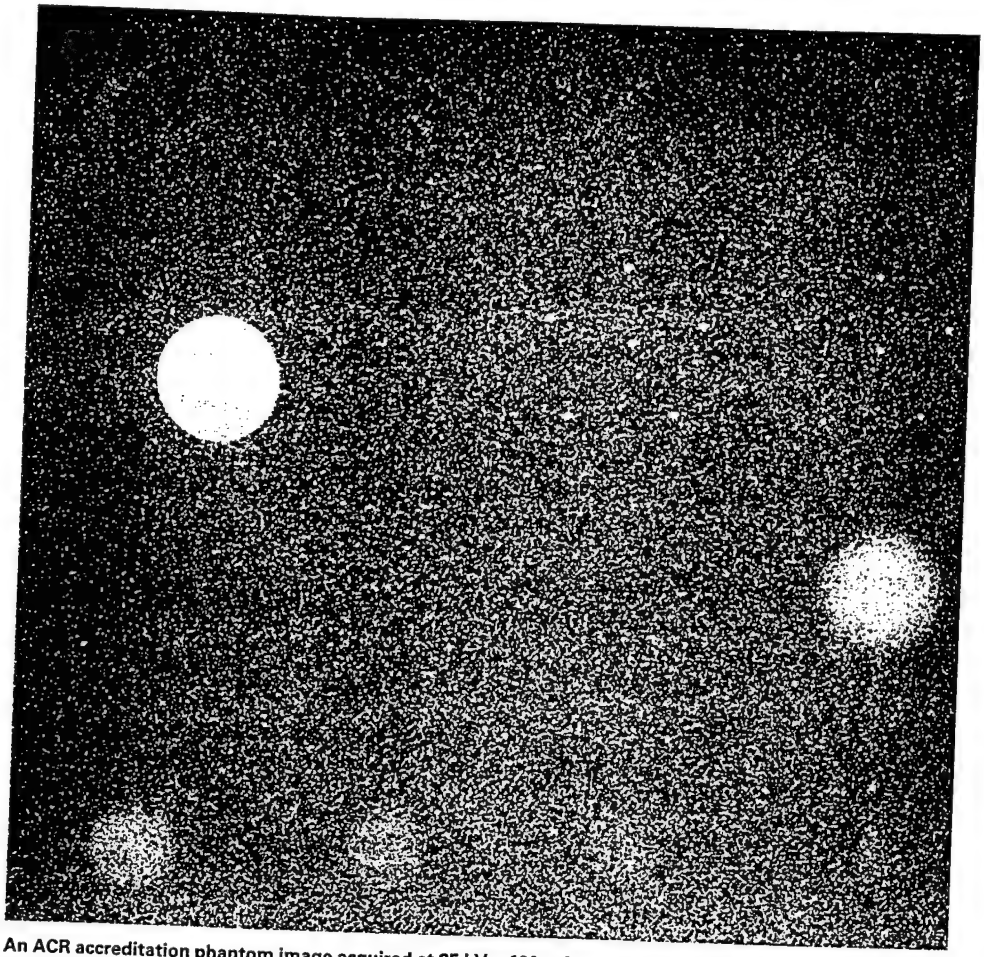

Fig 10. An ACR accreditation phantom image acquired at 25 kVp, 100 mAs, *Mo/Mo* target/filter combination, 660 mm SID, with antiscatter grid corresponding to a mean glandular dose of 1.23 mGy. The process of transferring the image onto a paper print resulted in a slight degradation, but the hard copy (printed film) and soft copy (CRT display) images indicated the capability of resolving 5 fibers, 4 speck groups, and 5 masses.

bleeding effects were observed in the SMPTE pattern image. The flat-field uniformity study conducted for all 3 target-filter combinations did not display artifacts like gridlines, horizontal or vertical streaks, or bad pixels. Wire mesh screen images indicated uniform resolution and no geometric distortion.

The most common and regularly used measure of image quality has been the scoring of ACR mammographic accreditation phantom. Also, a compelling argument has been presented with respect to the requirement of improved performance to justify the increased costs.[20] Based on these factors and the improved physical characteristics observed with such digital mammographic systems,[26] a logical requirement for these full-field digital mammographic systems might be the need to perform better than screen-film systems. We were able to detect regularly 4 speck groups and, probably more significantly, 5 masses, which denotes performance that exceeds that expected of screen-film systems. A practical quality assurance procedure should probably include the scoring of the ACR accreditation phantom on a weekly basis. In addition, because the ability to achieve consistent imaging performance would depend on the flat-field uniformity, performing gain, offset, and flat field calibrations on a weekly basis might be appropriate.

The full-breast digital mammographic imager did not exhibit any appreciable artifact or structured noise. The flat panel imager showed good

stability as observed by the conversion factor measurements. Although variations in detector signal response, linearity, sensitivity, and SNR were observed, interim results from a multicenter clinical trial do not show a significant difference between film-screen and full-field digital mammography in sensitivity for breast cancer detection.[34] The flat panel imager also showed a remarkable capability to image fibrils, speck groups, and masses in the ACR phantom. Clinical images show equally favorable results in visualizing soft tissue anatomy and calcifications.[34-36] Although the phantom and clinical images are very encouraging, the clinical efficacy in terms of sensitivity and specificity is the subject of a different investigation, which is currently in progress.[34]

ACKNOWLEDGMENT

The authors thank John Sandrik, PhD, GE Medical Systems for useful technical discussions.

REFERENCES

1. Bick U, Giger ML, Schmidt RA, et al: Density correction of peripheral breast tissue on digital mammograms. Radiographics 16:1403-1411, 1996

2. Kupinski MA, Giger ML: Automated seeded lesion segmentation on digital mammograms. IEEE Trans Med Imaging 17:510-517, 1998

3. Buchbinder SS, Leichter IS, Bamberger PN, et al: Analysis of clustered microcalcifications by using a single numeric classifier extracted from mammographic digital images. Acad Radiol 5:779-784, 1998

4. Anastasio MA, Yoshida H, Nagel R, et al: A genetic algorithm-based method for optimizing the performance of a computer-aided diagnosis scheme for detection of clustered microcalcifications in mammograms. Med Phys 25:1613-1620, 1998

5. Wei D, Chan HP, Petrick N, et al: False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis. Med Phys 24:903-914, 1997

6. Yoshida H, Doi K, Nishikawa RM, et al: An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms. Acad Radiol 3:621-627, 1996

7. Kallergi M, Carney GM, Gaviria J: Evaluating the performance of detection algorithms in digital mammography. Med Phys 26:267-275, 1999

8. Niklason LT, Christian BT, Niklason LE, et al: Digital tomosynthesis in breast imaging. Radiology 205:399-406, 1997

9. Ruttimann UE, Groenhuis RAJ, Webber RL: Restoration of digital multiplane tomosynthesis by a constrained iteration method. IEEE Trans Med Imaging MI-3:141-148, 1984

10. Suryanarayanan S, Karellas A, Vedantham S, et al: Tomosynthesis reconstruction methods for digital mammography. Biomedical imaging symposium: Visualizing the future of biology and medicine, National Institute of Health, June 25-26, Bethesda, MD, 1999

11. Glick SJ, Karellas A, Vedantham S, et al: Iterative reconstruction for limited angle tomographic digital mammography. Radiology 209P:159, 1998

12. Flanagan FL, Murray JG, Gilligan P, et al: Digital subtraction in Gd-DTPA enhanced imaging of the breast. Clin Radiol 50:848-854, 1995

13. Chakraborty DP, Barnes GT: An energy sensitive cassette for dual-energy mammography. Med Phys 16:7-13, 1989

14. Asaga T, Chiyasu S, Mastuda S, et al: Breast imaging: Dual-energy projection radiography with digital radiography. Radiology 164:869-870, 1987

15. Holdsworth DW, Gerson RK, Fenster A: A time-delay integration charge-coupled device camera for slot-scanned digital radiography. Med Phys 17:876-886, 1990

16. Nishikawa RM, Mawdsley GE, Fenster A, et al: Scanned-projection digital mammography. Med Phys 14:717-727, 1987

17. Karellas A, Harris LJ, Liu H, et al: Charge-coupled device detector: Performance considerations and potential for small-field mammographic imaging applications. Med Phys 19:1015-1023, 1992

18. Hejazi S, Trauernicht DP: System considerations in CCD-based x-ray imaging for digital chest radiography and digital mammography. Med Phys 24:287-297, 1997

19. Vedantham S, Levis I, Karellas A, et al: Characterization of a clinical prototype small-format CCD-based cassette for digital mammography. Radiology 209P:160, 1998

20. Kimme-Smith C, Lewis C, Beifuss M, et al: Establishing minimum performance standards, calibration intervals, and optimal exposure values for a whole breast digital mammography unit. Med Phys 25:2410-2416, 1998

21. Siewerdsen JH, Antonuk LE, El-Mohri Y, et al: Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology. Med Phys 24:71-89, 1997

22. Antonuk LE, El-Mohri Y, Jee K, et al: Performance evaluation of a large area, 97 μm pitch: Indirect detection active matrix flat-panel imager (AMFPI) for radiography and fluoroscopy. Radiology 209P:357, 1998

23. Antonuk LE, El-Mohri Y, Jee K, et al: Performance limits of high resolution large area active matrix flat-panel imagers (AMFPIs). Radiology 209P:581, 1998

24. Antonuk LE, Jee K, El-Mohri Y, et al: Strategies to significantly enhance performance of active matrix flat-panel imagers (AMFPIs). Radiology 209P:358, 1998

25. Rougeot HM, Opsahl-Ong B, Castleberry DE, et al: Performance evaluation of a flat-panel filmless full-field digital mammography system. Radiology 201P:190, 1996

26. Vedantham S, Karellas A, Suryanarayanan S, et al: Full breast digital mammographic imaging with an amorphous silicon-based flat panel detector: Physical characteristics of a clinical prototype. Med Phys 27:558-567, 2000

27. Karellas A, Vedantham S, Levis I, et al: Evaluation of a full-field clinical prototype flat panel imager for digital mammography. Radiology 209P:159, 1998

28. Neitzel U, Maack I, Gunther-Kohfahl S: Image quality of

a digital chest radiography system based on a selenium detector. Med Phys 21:509-516, 1994

29. Zhao W, Rowlands JA: Digital radiology using active matrix readout of amorphous selenium: Theoretical analysis of detective quantum efficiency. Med Phys 24:1819-1833, 1997

30. Garverick SL, Skrenes L, Baertsch RD: A 32-channel charge readout IC for programmable, nonlinear quantization of multichannel detector data. IEEE J of Solid-State Circuits 30:533-541, 1995

31. Fujita H, Tsai DY, Itoh T, et al: A simple method for determining the modulation transfer function in digital radiography. IEEE Trans Med Imaging MI-11:34-39, 1992

32. American College of Radiology: Mammography Quality Control Manual, Committee on Quality Assurance in Mammography, American College of Radiology, 1999

33. Nawfel RD, Chan KH, Wagenaar DJ, et al: Evaluation of video gray-scale display. Med Phys 19:561-567, 1992

34. Lewin JM, Hendrick RE, D'Orsi CJ, et al: Clinical evaluation of a full-field digital mammography prototype for cancer detection in a screening setting—Work in progress. Radiology 209P:238, 1998

35. Moss L, D'Orsi CJ, Karellas A, et al: Initial experience with a high resolution full field digital mammographic system. J Digit Imaging 11:110, 1998

36. Moore RH, Kopans DB, Niklason LT, et al: Initial clinical experience with full-field digital mammography. Radiology 205P:274, 1997

Full breast digital mammography with an amorphous silicon-based flat panel detector: Physical characteristics of a clinical prototype

Srinivasan Vedantham, Andrew Karellas,[a] and Sankararaman Suryanarayanan
Department of Radiology, UMass Memorial Health Care, University of Massachusetts Medical School, Worcester, Massachusetts 01655

Douglas Albagli, Sung Han, Eric J. Tkaczyk, Cynthia E. Landberg, and Beale Opsahl-Ong
G. E. Corporate Research and Development, Niskayuna, New York 12309

Paul R. Granfors
G. E. Medical Systems, Milwaukee, Wisconsin 53201

Ilias Levis and Carl J. D'Orsi
Department of Radiology, UMass Memorial Health Care, University of Massachusetts Medical School, Worcester, Massachusetts 01655

R. Edward Hendrick[b]
Radiology Department, University of Colorado Health Science Center, Denver, Colorado 80262

(Received 28 June 1999; accepted for publication 28 December 1999)

The physical characteristics of a clinical prototype amorphous silicon-based flat panel imager for full-breast digital mammography have been investigated. The imager employs a thin thallium doped CsI scintillator on an amorphous silicon matrix of detector elements with a pixel pitch of 100 μm. Objective criteria such as modulation transfer function (MTF), noise power spectrum, detective quantum efficiency (DQE), and noise equivalent quanta were employed for this evaluation. The presampling MTF was found to be 0.73, 0.42, and 0.28 at 2, 4, and 5 cycles/mm, respectively. The measured DQE of the current prototype utilizing a 28 kVp, Mo–Mo spectrum beam hardened with 4.5 cm Lucite is ~55% at close to zero spatial frequency at an exposure of 32.8 mR, and decreases to ~40% at a low exposure of 1.3 mR. Detector element nonuniformity and electronic gain variations were not significant after appropriate calibration and software corrections. The response of the imager was linear and did not exhibit signal saturation under tested exposure conditions. © *2000 American Association of Physicists in Medicine.* [S0094-2405(00)01803-4]

Key words: breast imaging, digital mammography, physics, image quality, detective quantum efficiency (DQE)

I. INTRODUCTION

The physical aspects of mammography have been the subject of many investigations which have addressed basic imaging characteristics such as x-ray scatter,[1–4] x-ray tube focal spot effects,[5] and x-ray spectra.[6,7] This knowledge has served as the basis for many technical improvements and regulatory standards of performance.[8]

Though film-screen mammography is currently the standard in breast imaging, it has well-known limitations with regard to dynamic range, contrast, and lack of convenient options for postprocessing of images. It is apparent that electronic detection has the theoretical capability of overcoming certain fundamental limitations of film-screen systems. The potential advantages of electronic detection include high detection efficiency, high dynamic range, capability of contrast enhancement,[9] and postprocessing capabilities including computer-aided diagnosis.[10–15] Further, direct electronic acquisition enables the exploration of novel imaging techniques such as tomosynthesis,[16,17] dual-energy mammography,[18,19] and digital subtraction imaging.[20] In the past, investigators have used different modes of electronic detection technology to gain insight into electronic mammography, commonly referred to as digital mammography.[21] Early evaluations have used image intensifiers and subsequently slot-scanned systems[22,23] with charge-coupled devices (CCDs) and CCDs with fiberoptic tapers.[24] Development of an electronic detector to cover the entire breast presents a formidable technical challenge. Currently, digital mammography is limited to small field devices for stereotactic localization, core biopsy, and spot compression views.[24,25] It is now feasible to manufacture large flat panel monolithic arrays of amorphous silicon photodiodes coupled to thin-film transistors on a glass substrate. These arrays utilize a scintillator as the primary detection layer to convert x rays to light, which is subsequently detected by the photosensing silicon elements. Several studies characterizing amorphous silicon[26–30] and amorphous selenium[31,32] based imagers for chest radiography and other applications have been reported in the recent past. However, detailed experimental characterization of amorphous silicon based flat panel imagers under realistic mammographic conditions have not been reported in the past.

This study characterizes the image quality parameters of

TABLE I. Amorphous silicon-based flat panel detector specifications.

Flat panel image area	18 cm×23 cm
Pixel matrix	1800×2304
Pixel size	100 μm
Scintillator	CsI:Tl

an amorphous silicon-based clinical prototype flat panel imager (GE Medical Systems, Milwaukee, WI) presently undergoing technical and clinical evaluation at the University of Massachusetts Medical School and the University of Colorado Health Sciences Center.

II. METHODS AND MATERIALS

The full-breast digital mammography imager characterized in this study is composed of a thallium-doped CsI scintillator and an amorphous silicon photodiode array and incorporates special-purpose readout electronics. Light created from the interaction of x-ray photons in the scintillator travels down the columnar crystalline structure of the scintillator, which is in contact with a two-dimensional array of amorphous silicon photodiodes and thin-film transistors. Light exiting from the scintillator is detected by the monolithic thin film flat panel array, which consists of a matrix of 1800 ×2304 detector elements that are 100 μm in pitch. The specifications of the mammographic flat panel imager are presented in Table I. Each detector element (pixel) in the array is an individually addressable light detector. The electrical signals of all pixels are individually read out and digitized to 16 bit digital values in 300 ms by special-purpose low-noise electronics[33] which are located inside the image receptor assembly. The schematic of the detector is shown in Fig. 1. The imager is integrated into a prototype digital mammography system based on a multipulse high frequency x-ray generator (Senographe DMR, GE Medical Systems, Milwaukee, WI). This system uses a selectable dual track target, either molybdenum (Mo) or rhodium (Rh) with selectable filtration of Mo or Rh. All measurements were performed at 28 kVp with a Mo/Mo target/filter combination. This particular technique was chosen as it was found to be

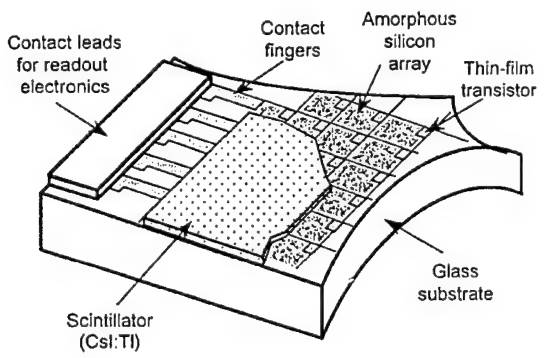

FIG. 1. Schematic of the amorphous silicon detector array.

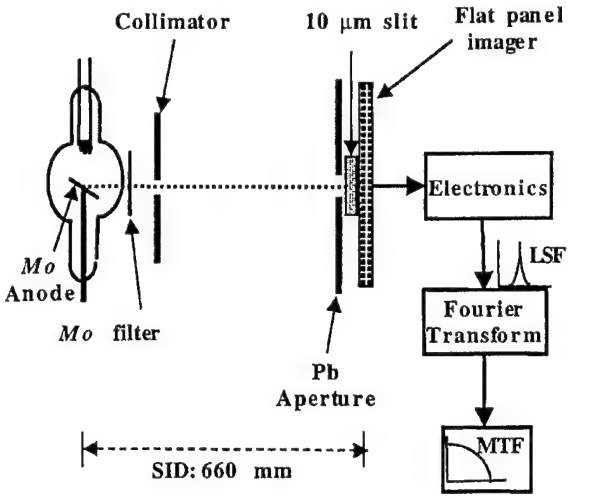

FIG. 2. Experimental setup for MTF measurement. The area surrounding the 10 μm slit was covered with Pb (0.5 cm thick).

the median exposure technique used in a random sample of 100 breast exams from a population of 1400 patients performed with this flat panel imager.

A. Presampling modulation transfer function measurement

The presampling modulation transfer function (MTF) was measured according to the technique described by Fujita *et al.*[34] The experimental procedure for measuring the same has also been described in detail by Dobbins *et al.*[35] The effects of undersampling have also been described in detail by Dobbins.[36] The experimental setup is shown in Fig. 2. An image of a 10-mm-long, 10 μm ($\pm$1 μm) slit made of 1.5-mm-thick tantalum placed at a slight angle (less than 4°) to the anode–cathode axis at the center of the detector was obtained. The area around the slit was covered with Pb (0.5 cm thick). The slit was placed about 5.5 mm (due to thickness of the breast support plate and the slit housing) from the surface of the imager. Since the magnification of the slit was about 1.0083, there was no appreciable spreading of the line spread function (LSF) due to focal spot blurring. The exposure technique was adjusted to ensure that the tails of the dark image subtracted LSF obtained had no significant electronic noise. The appropriate technique found to be 28 kVp, 160 mAs was used. The source-to-image distance was maintained at 660 mm during the study. The image of the slit was obtained without the antiscatter grid in place. The slit image obtained was corrected for variations along the edge of the slit. This was accomplished by normalizing the signal values along the horizontal direction (perpendicular to the anode–cathode axis) by dividing each pixel value by the sum of the pixel values in that particular row as illustrated in Fig. 3. This normalization method assumes that the slit width is approximately constant over the length used for obtaining the finely sampled LSF and that the signal spreading is approximately equal along each line of data. The validity of these assumptions was verified by calculating the MTF from sev-

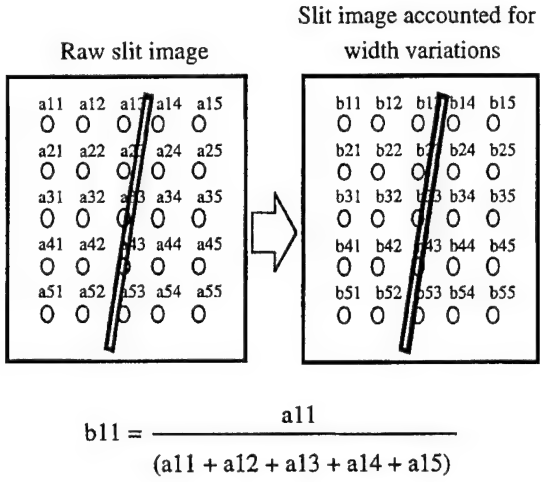

FIG. 3. Illustration of slit image correction for variations in slit width.

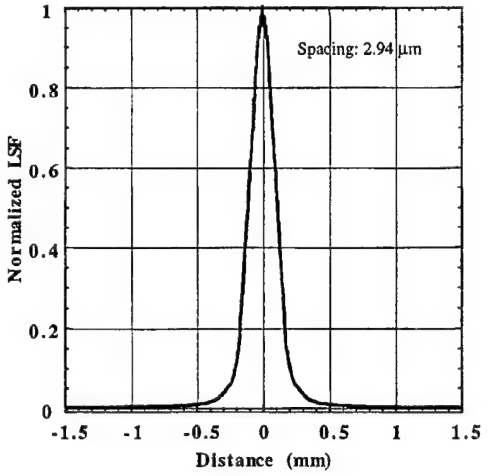

FIG. 5. Finely sampled LSF.

eral locations along the central region of the slit, and the MTF was found to vary by less than 1%. Before performing this normalization care was taken to avoid loss of information due to truncation by converting the pixel intensity values to 32 bit floating point numbers. The pixel amplitudes along the column or vertical direction (along the anode–cathode axis) were plotted as shown in Fig. 4. This provided the adequate number of individual LSFs needed to obtain a finely sampled LSF. Since each pixel represented a sample of the LSF at a distance equal to the distance between the center of the slit and the pixel center, the finely sampled LSF was obtained by plotting the pixel intensity from the center of the slit. The finely sampled LSF was synthesized by using 34 individual LSFs and normalized to a peak value of one (Fig. 5). The Fourier transform (FT) of the finely sampled LSF was performed and the resultant FT was deconvolved of the finite dimension of the slit by dividing the resultant FT by a sinc function in the frequency domain to provide the presampling MTF. The presampling MTF was measured along both the horizontal (perpendicular to the anode–cathode axis) and vertical (along the direction of the anode–cathode axis) directions.

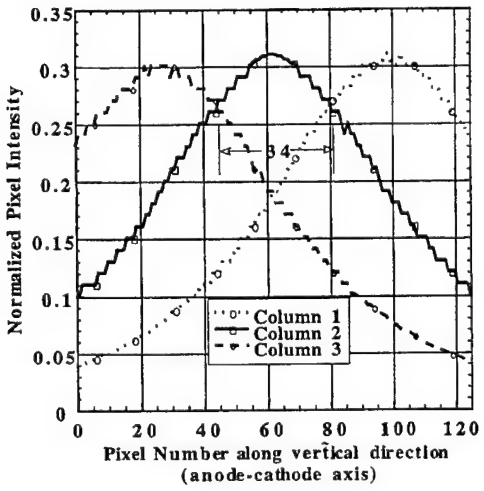

FIG. 4. The pixel amplitudes along the anode–cathode axis used for determining the number of rows of data needed to obtain a finely sampled LSF.

B. Noise power spectrum measurement

There are many inherent difficulties in measuring the noise power spectrum (NPS) of digital systems.[35–38] Computing the two-dimensional (2D) NPS is important to study the presence or absence of any off-axis noise peaks. Since the computation time of computers is no longer a constraint,[35] computing the entire 2D NPS and estimating the one-dimensional (1D) NPS from the 2D NPS was used. The 1D NPS was estimated from the 2D NPS using the technique described by Dobbins *et al.*[35] This technique utilizes a thick cut parallel to and immediately adjacent to the axes for estimating the 1D NPS. We used the data in a thick slice comprised of eight lines on either side of both the axes (excluding the axes). For each data value at (u,v) in this thick slice, the frequency value was computed as $\sqrt{u^2+v^2}$ for the 1D NPS estimate. The assumptions for utilizing this technique for estimating the 1D NPS are that the 2D NPS exhibit moderate radial symmetry and that the noise data are nominally uniform within the small annuli of spatial frequencies used for regrouping the noise data.

The next major difficulty was to determine the finite window of the noise data required to provide adequate resolution for proper representation of the NPS without the finite window overtly affecting the NPS estimate. Since the measured NPS is produced by convolving the "true" NPS with the sinc^2 function in the frequency domain, due to the finite window of the noise data, the choice of region-of-interest (ROI) size has to be considered carefully. We estimated the NPS using ROI sizes of 512×512, 256×256, 128×128, and 64×64, and determined the 256×256 ROI to be the smallest ROI required for proper representation of the NPS with minimum spectral distortion (spectral deviation between 512×512 ROI and 256×256 ROI was less than 5% over the

entire frequency range and the spectral deviation increased with smaller ROI sizes). Hence, the 256×256 ROI was utilized for NPS estimations in the entire study.

The other difficulty was to determine the number of NPS realizations needed to be averaged in order to obtain a smooth and accurate curve depicting the noise spectrum. Ideally, we would need a large number of NPS realizations so that they can be averaged to obtain a smooth spectrum. We considered 10, 15, 20, 30, and 50 NPS realizations and found that the ensemble average of 15 NPS realizations taken from the same location through 15 images was sufficient to accurately characterize the NPS of the system. We were able to achieve a smooth spectrum by averaging eight lines of data on either side of the axes.

Problems associated with background trends such as from the heel effect can corrupt the noise spectrum and provide artificially inflated values[35,38] along the axes. However, techniques for suppression of such background trends have been described by various authors.[35,38] We surface (ramp) fitted each ROI and subtracted these background trends. Though this method was successful in suppressing these background trends, it did not completely eliminate them. Hence, we avoided using data values directly on the axes, as they were not representative in amplitude of the rest of the 2D NPS in the vicinity of the axes.

In order to measure the noise power spectra of the detector the detector has to be linear and shift invariant.[39] The linear response and sensitivity of the system was measured by averaging the pixel intensity over a 256×256 ROI centered at the 4 cm from the chest wall edge of the detector at various exposure levels. All images for the noise power spectral estimate used for calculation of detective quantum efficiency (DQE) were dark subtracted [Eq. (1)] and flat field corrected [Eq. (2)] resulting in a nominally uniform image,

$$\text{dark subtracted}_i(x,y) = \text{flood}_i(x,y) - \text{dark}_i(x,y), \tag{1}$$

$$\text{flat field}_i(x,y) = \frac{\text{dark subtracted}_i(x,y)}{(1/n)\sum_{i=1}^{n}\text{dark subtracted}_i(x,y)} \times \frac{1}{m^2}\sum_{y=1}^{m}\sum_{x=1}^{m}\left[\frac{1}{n}\sum_{i=1}^{n}\text{dark subtracted}_i(x,y)\right], \tag{2}$$

where $\text{flood}_i(x,y)$ and $\text{dark}_i(x,y)$ represent the flood and dark ROIs, respectively;
$(1/n)\Sigma_{i=1}^{n}$ dark subtracted$_i(x,y)$ is the average of the dark subtracted ROIs; $1/m^2\Sigma_{v=1}^{m}\Sigma_{u=1}^{m}[(1/n)\Sigma_{i=1}^{n}\text{dark subtracted}_i \times(x,y)]$, is the mean of the average of the dark subtracted ROIs; and, in our case, $m=256$ and $n=15$. The ROIs (256×256) used for the NPS analysis were taken from the same location (centered at 4 cm from the chest wall edge of the detector) from multiple (15) images. Though the detector might not to be completely shift invariant, the process of flat field correcting and using the same ROI from multiple images for NPS analysis allows for the reasonable assumption of the "shift-invariant" property of the system.

The noise power spectra were determined at four exposure levels and were obtained with 4.5-cm-thick Lucite in the x-ray beam path. This thickness of Lucite was used as it was found to be the median thickness range (4.5–4.99 cm) of the compressed breast from a random sample of 100 breast exams obtained from a population of 1400 patients. The antiscatter grid was not used while obtaining the images as it might provide a possible noise source, which might corrupt the measurement. In order to minimize scattered radiation affecting the measurement due to the removal of the antiscatter grid, the 4.5-cm-thick Lucite block was mounted on to the tube housing. In addition, the x-ray beam was collimated both at the tube port and at the surface of the detector using Pb (0.5 cm) so that only a 4 cm×4 cm area of the detector was irradiated. This enabled us to obtain our objective of achieving a realistic clinical spectrum without the measurement being affected by either excessive scattered radiation or the presence of structure from an antiscatter grid. The setup for NPS measurement is shown in Fig. 6. Fifteen dark image subtracted, flat field corrected, 256×256 ROIs were acquired as described previously. Before performing dark image subtraction and flat field correction, care was taken to avoid information loss due to truncation by converting the pixel intensity values to 32 bit floating point numbers from the original 16 bit digital values. A surface fit (like a ramp) to suppress background trends like heel effect was performed on each ROI. The ensemble average of the squares of the magnitude of these 15 Fourier transformed 256×256 ROIs scaled as shown in Eq. (3) provided the 2D raw noise power spectrum, $\text{NPS}_{\text{raw}}(u,v)$.[35]

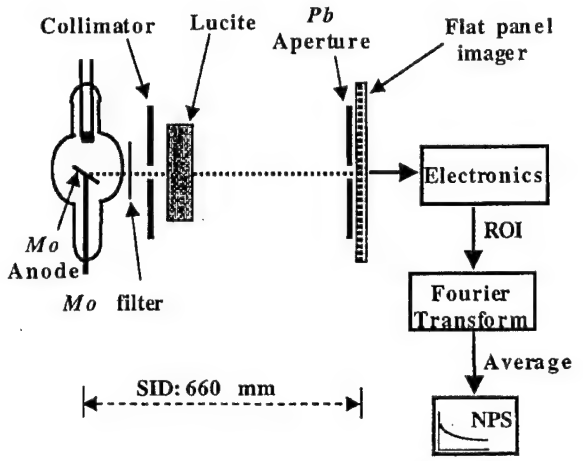

FIG. 6. Experimental setup for NPS measurement where a 4 cm×4 cm area of the detector centered at 4 cm from the chest wall edge was irradiated. Lead collimation at the tube port and at the detector surface reduced excessive scatter.

The $\text{NPS}_{\text{raw}}(u,v)$ was obtained by

$$\text{NPS}_{\text{raw}}(u,v) = \frac{\langle|\text{FT}[\text{flat field}(x,y)]|^2\rangle}{N_x N_y}\Delta_x\Delta_y, \tag{3}$$

where $\langle|\text{FT}[\text{flat field}(x,y)]|^2\rangle$ represents the ensemble average of the squares of the magnitude of the Fourier transformed 256×256 ROIs, N_x and N_y are the number of ele-

ments in the x and y directions, respectively (which are equal and is 256 in this case), and Δ_x and Δ_y are the pixel pitch in x and y directions, respectively (which are equal and is 100 μm with this imager).

To compute noise equivalent quanta (NEQ) and DQE a 1D NPS curve was required. This was achieved by using the data in a thick slice comprised of eight lines on either side of both the u and v axes (excluding the axes). For each data value at (u,v) in this thick slice, the frequency value was computed as $\sqrt{u^2+v^2}$ for the 1D NPS estimate. The final 1D NPS at each exposure level is the average of 8 (lines) $\times 2$(sides) $\times 256$ data points (=4096 data values) grouped into frequency bins 0.04 mm^{-1}. The 1D $\text{NPS}_{\text{normalized}}(f)$ to be used for the DQE calculations was obtained by scaling the 1D $\text{NPS}_{\text{raw}}(f)$ for the mean signal by

$$\text{NPS}_{\text{normalized}}(f)=\frac{\text{NPS}_{\text{raw}}(f)}{(\text{mean signal of } 256\times 256\,\text{ROI})^2}. \tag{4}$$

The mean signal of the 256×256 ROI is expressed in digital values.

The electronic noise present in the system was also estimated. The entire detector was covered with Pb (2 cm) and 15 images were acquired using the minimum possible exposure technique. The 2D $\text{NPS}_{\text{electronic}}(u,v)$ was estimated as per Eq. (3) at this minimum possible exposure technique with Pb, and the 1D $\text{NPS}_{\text{electronic}}(f)$ estimated by using a thick slice as described earlier. From this measurement, the noise contribution due to the x rays, $\text{NPS}_{\text{x ray}}(f)$ was calculated at each exposure level as per Eq. (5), where $\text{NPS}_{\text{raw}}(f)$ is the raw NPS estimated as per Eq. (3) and $\text{NPS}_{\text{electronic}}(f)$ is the electronic noise of the system. The x-ray component of $\text{NPS}_{\text{raw}}(f)$ was computed as per Eq. (6)

$$\text{NPS}_{\text{x ray}}(f)=\text{NPS}_{\text{raw}}(f)-\text{NPS}_{\text{electronic}}(f), \tag{5}$$

$$\text{x-ray component of } \text{NPS}_{\text{raw}}(f)=\frac{\text{NPS}_{\text{x ray}}(f)}{\text{NPS}_{\text{raw}}(f)}\times 100\%. \tag{6}$$

In order to study the structured noise component or the presence of any varying nonstochastic noise, the 2D $\text{NPS}_{\text{subtracted}}(u,v)$ was estimated as per Eqs. (7) and (8). Background suppression (ramp fit) was not performed for estimation of $\text{NPS}_{\text{subtracted}}(u,v)$. The 1D $\text{NPS}_{\text{subtracted}}(f)$ was obtained by using a thick slice of eight lines of data on either side of the axes as described earlier,

$$\text{residual}_i(x,y)=[\text{flood}_i(x,y)-\text{dark}_i(x,y)]-\frac{1}{n}\sum_{i=1}^{n}\text{flat field}_i(x,y), \tag{7}$$

$$\text{NPS}_{\text{subtracted}}(u,v)=\frac{\langle|\text{FT}(\text{residual}(x,y))|^2\rangle}{(\text{mean signal of } 256\times 256\,\text{ROI})^2 N_x N_y}\Delta_x\Delta_y. \tag{8}$$

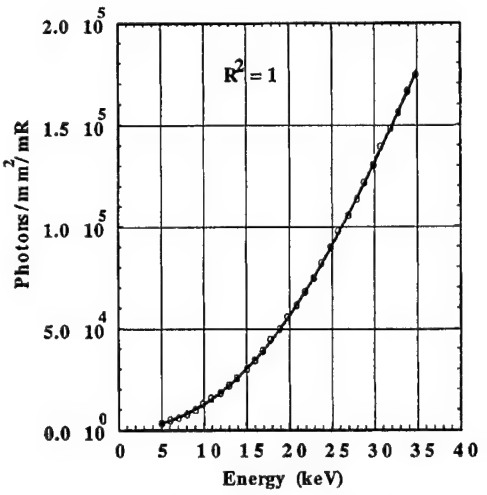

FIG. 7. Curve fitted x-ray photon fluence per mR between the energy range of 5 and 35 keV obtained from published values.

C. NEQ and DQE measurement

The NEQ was computed as[35]

$$\text{NEQ}(f)=\frac{\text{MTF}^2(f)}{\text{NPS}_{\text{normalized}}(f)}. \tag{9}$$

The NEQ of the system was computed for the four exposure levels. For the purpose of calculating the DQE of the digital imager, Eqs. (10) and (11) were used:[35]

$$\text{DQE}(f)=\frac{\text{MTF}^2(f)}{\text{NPS}_{\text{normalized}}(f)q}. \tag{10}$$

and hence

$$\text{DQE}(f)=\frac{\text{NEQ}(f)}{q}, \tag{11}$$

where $\text{MTF}(f)$ is the modulation transfer function of the system; $\text{NPS}_{\text{normalized}}(f)$ is the normalized noise power spectrum of the imaging system; q is the number of x-ray photons incident on the detector per unit area; $\text{NEQ}(f)$ is the noise equivalent quanta of the imaging system and f is the spatial frequency. The only factor that needs to be determined is q.

Determination of q. Determination of q was done in three stages. First, the x-ray photon fluence per mR was curve fitted between the energy range of 5 to 35 keV from already published values[40] and is shown in Fig. 7. The photon fluence per mR, $Y(e)$, at energy (e) is best described by the polynomial:

$$Y(e)=2.2128+33.514e+89.23e^2+3.0588e^3-0.0239e^4-0.0006e^5-3\times 10^{-7}e^6. \tag{12}$$

The x-ray spectral distribution, $q(e)$, was characterized by averaging 15 spectra obtained using a cadmium zinc telluride (CZT) based high resolution spectrometer (XR-100T-CZT, Amptek, Inc., USA). The x-ray spectrum was corrected for dead time losses and pile-up.[41] Correction for the spectrometer energy response was not needed as the energy ab-

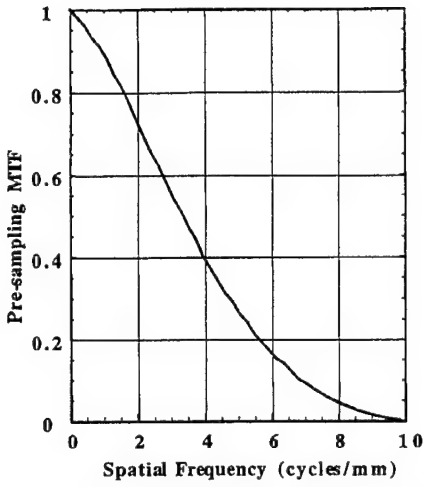

FIG. 8. The presampling MTF of the full field flat panel a:Si imager.

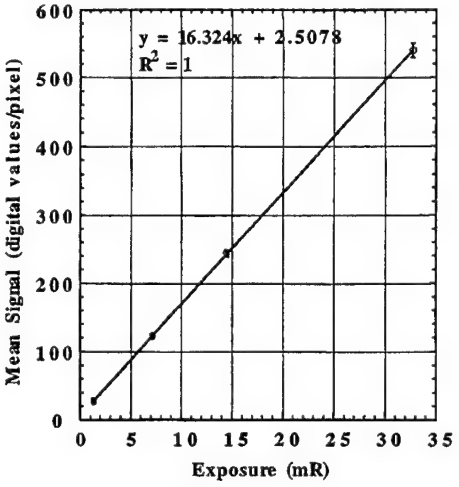

FIG. 9. Linearity of the system. The data points represent the mean intensity and the error bars represent the standard deviation from this mean value.

sorption efficiency of the 3-mm-thick CZT spectrometer is more than 99.9% for the energy range (5–35 keV) of the incident spectrum. The exposure (X) on the surface of the detector was measured under the same conditions as during the NPS measurement with a calibrated mammographic ionization chamber connected to MDH 1515 (RadCal Corp., USA) dosimeter. The precision at each exposure level was improved by averaging five measurements. The total number of photons incident per unit area of the detector at each exposure level was calculated as per Eq. (13). With the knowledge of q, the DQE(f) was calculated,

$$q = X\frac{\int q(e)Y(e)de}{\int q(e)de}. \tag{13}$$

III. RESULTS AND DISCUSSION

A. Presampling MTF

The measured presampling MTF is shown in Fig. 8. The presampling MTF measured both along the vertical and horizontal directions were identical. The presampling MTF was found to be 0.73, 0.42 and 0.28 at 2, 4, and 5 cycles/mm, respectively. Although the MTF of an imaging system is an important objective measure of the spatial resolution, this parameter alone may not be predictive of the overall performance of the system. Other metrics such as DQE as a function of the spatial frequency provide additional insight.

B. Noise power spectra

The linearity of the system was measured and is shown in Fig. 9. From the linearity measurements the sensitivity of the system was found to be 16.324 digital values/mR/pixel. The 2D NPS obtained at 1.3, 7.1, 14.5, and 32.8 mR are shown in Figs. 10(a), 10(b), 10(c), and 10(d), respectively. The noise power at the intersection of the u and v axes are much higher in magnitude and hence this point has been blanked for display purposes. The images are displayed in a black and white scheme where the transition point is set at the midpoint of the minimum and maximum of the 2D NPS images. The 2D NPS does not show the presence of any off-axis noise peaks. The 1D NPS_{raw} at four exposure levels of 1.3, 7.1, 14.5, and 32.8 mR are shown in Fig. 11. The electronic noise present in the system is also shown in Fig. 11. The 1D NPS_{raw} demonstrates an increase in noise with increasing exposure as the photon noise increases with increasing exposure. The integral of the NPS at each exposure was confirmed to be identical to the rms variance of the 256×256 ROI. Figure 12 shows the x-ray component of the total NPS calculated as

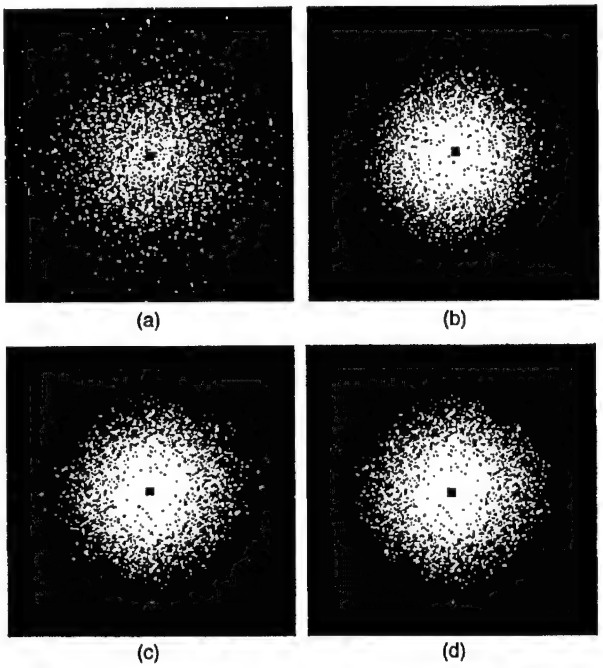

FIG. 10. The 2D NPS obtained at 1.3, 7.1, 14.5, and 32.8 mR are shown in (a), (b), (c), and (d), respectively. The intersection of the axes has been masked for display purposes. The images are displayed in a black and white scheme, with the transition point set at the mean of the ROI.

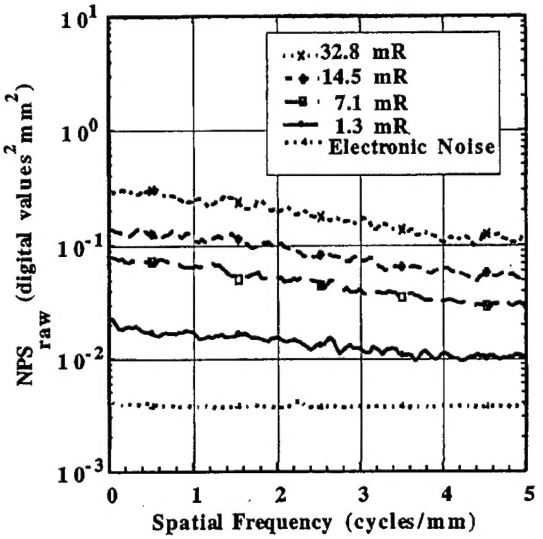

FIG. 11. The 1D noise power spectra (NPS_{raw}) at four exposure levels of 1.3, 7.1, 14.5, and 32.8 mR are shown. The electronic noise is also shown.

per Eq. (6) at the four exposure levels. Even at a low exposure of 1.3 mR, the x-ray component was dominant (greater than 60% of the total NPS at 5 cycles/mm and approximately 80% of the total NPS at ~0 cycle/mm). Figure 13 suggests that there is no appreciable structure noise or varying nonstochastic noise at exposures of 1.3 and 32.8 mR as the $NPS_{subtracted}$ and $NPS_{normalized}$ are identical.

C. NEQ and DQE

The NEQ of the system at four exposure levels are shown in Fig. 14. The Mo–Mo spectrum incident on the detector transmitted through 4.5 cm of Lucite and the breast support plate recorded with a high resolution spectrometer is shown in Fig. 15. From this spectral distribution and Fig. 7, the photon flux incident on the detector was determined to be 0.533×10^5 photons/mm^2/mR. The DQE of the system at four exposure levels is shown in Fig. 16. To demonstrate the exposure dependence of the DQE of the system, DQE (0.2 cycle/mm), DQE(1 cycle/mm), DQE(2 cycles/mm), DQE(3 cycles/mm), and DQE(5 cycles/mm) are plotted as a function of the incident exposure in Fig. 17. The plot indicates that the DQE of the system increases with increasing exposure, and reaches a constant value at about 15 mR. The lower values of DQE at low exposures are primarily due to the contribution of electronic noise in the system. The DQE (~0 cycle/mm) was found to be 0.4, 0.48, 0.54, and 0.55 at incident exposures of 1.3, 7.1, 14.5, and 32.8 mR, respectively.

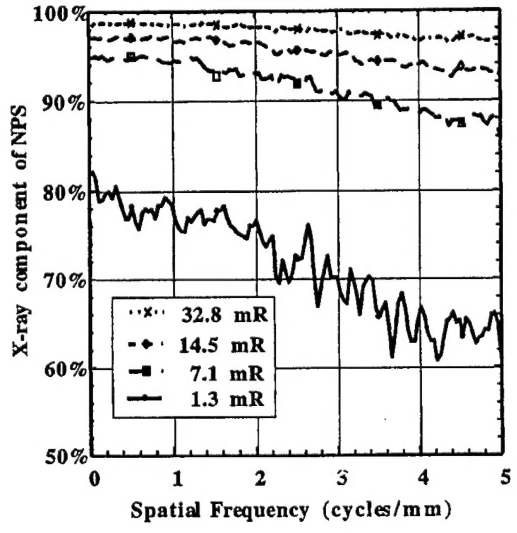

FIG. 12. The x-ray component of NPS_{raw} at four exposure levels of 1.3, 7.1, 14.5, and 32.8 mR are shown.

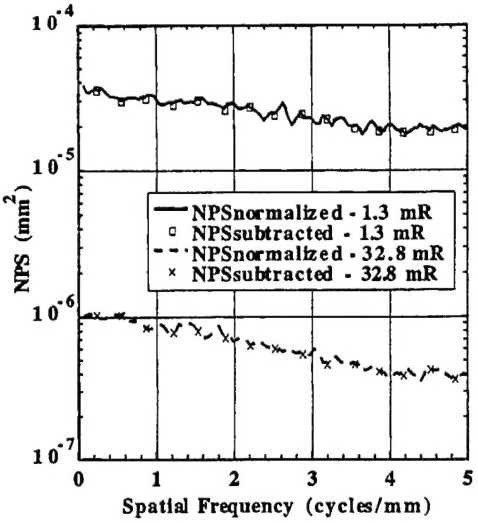

FIG. 13. The 1D $NPS_{normalized}$ and $NPS_{subtracted}$ obtained at 1.3 and 32.8 mR.

D. Discussion

Metrics such as MTF and DQE have been widely used to describe the performance characteristics of imaging systems.

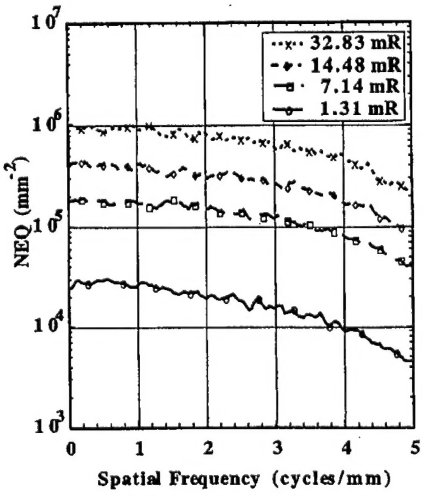

FIG. 14. The NEQ of the system at four exposure levels.

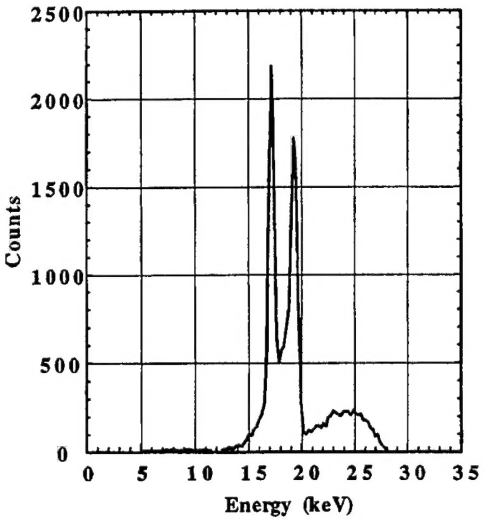

FIG. 15. The Mo–Mo spectra incident on the detector transmitted through 4.5 cm of Lucite and the breast support plate, recorded with a high resolution spectrometer for calculation of q.

A comparison of the flat-panel imager with other imaging systems such as screen-film systems to show the general trends could provide additional information as to the advantages and limitations of the flat-panel imager. Nishikawa and Yaffe[42] have evaluated various mammographic screen-film systems in the past. More recently, Bunch[43] has also evaluated the MTF and DQE of two widely used mammographic screen-film systems. Their results show a maximum DQB(~0) of 0.35 compared with 0.55 measured with the flat panel imager. The improved DQE of the flat panel imager at low and midfrequencies can be particularly advantageous in the imaging of low-contrast soft tissue lesions.[44,45] Their results also indicate that the spatial resolution is much higher

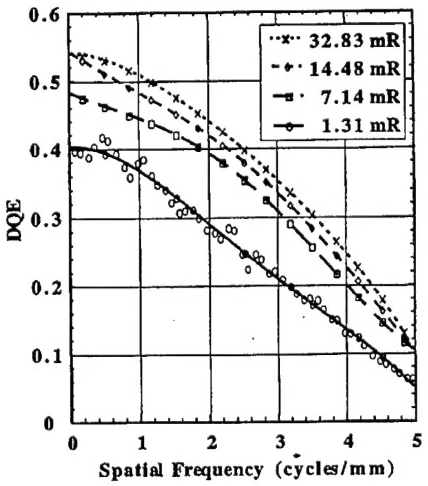

FIG. 16. The DQE of the system at four exposure levels. Data points are curve fitted with a sixth-order polynomial for clarity. To demonstrate the goodness of fit, data points at an exposure of 1.3 mR are shown.

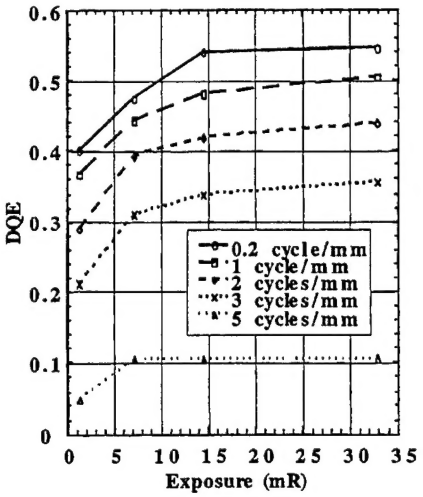

FIG. 17. DQE of the system plotted as a function of incident exposure.

with screen-film systems[42,43] compared to the flat-panel imager, but an increased film noise at high frequencies have also been observed.

Previous laboratory studies[23] in digital mammographic imaging using different technology have suggested that even with lower spatial resolution, lesion detectability, including microcalcifications can be improved by contrast enhancement of digital data. Prior work with this flat-panel imager has demonstrated a high dynamic range.[46] Clinical images with the current prototype demonstrate encouraging results for visualization of soft tissue anatomy and calcifications.[47–49] The clinical efficacy in terms of sensitivity and specificity is the subject of a different investigation, which is currently in progress.[47]

IV. CONCLUSIONS

A consistent set of image quality measurements was performed characterizing the full field amorphous silicon-based flat panel imager for mammographic applications. The flat panel imager did not exhibit any appreciable structured noise or varying nonstochastic noise component at the tested exposure levels. The response of the imager was linear and exhibited high sensitivity under tested exposure conditions. The flat panel imager demonstrated good dose efficiency within the tested exposure range.

ACKNOWLEDGMENTS

The authors would like to thank James T. Dobbins, Ph.D., Duke University Medical Center, Durham, NC27710 for useful scientific and technical discussions. The authors from GE Corporate Research and Development and GE Medical Systems acknowledge the invaluable assistance of many present and former colleagues. In particular, Gene Hilton, Ph.D., played a key role in the early development of the digital detector measurement technology. This work was supported in part by US Army Grant No DAMD17-96-C-6104 to the University of Colorado Health Sciences Center, and in part

by Grant No. R01CA59770 from the National Cancer Institute to the University of Massachusetts Medical School. The full-field digital mammography detector was developed independently by GE Corporate Research and Development with partial support from the National Cancer Institute Grant No. 5R01CA60183. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of NCI, NIH, or US Army.

[a]Author to whom correspondence should be addressed; electronic mail: andrew.karellas@umassmed.edu

[b]Present address: Breast Imaging Section, Northwestern University, Chicago, Illinois 60611.

[1] D. P. Chakraborty, "The effect of the antiscatter grid on full-field digital mammography phantom images," J. Digit. Imaging **12**, 12–22 (1999).

[2] Z. Jing, W. Huda, and J. K. Walker, "Scattered radiation in scanning slot mammography," Med. Phys. **25**, 1111–1117 (1998).

[3] J. Persliden and G. A. Carlsson, "Scatter rejection by air gaps in diagnostic radiology. Calculations using a Monte Carlo collision density method and consideration of molecular interference in coherent scattering," Phys. Med. Biol. **42**, 155–175 (1997).

[4] R. Fahrig, J. G. Mainprize, N. Robert, A. Rogers, and M. J. Yaffe, "Performance of glass fiber antiscatter devices at mammographic energies," Med. Phys. **21**, 1277–1282 (1994).

[5] H. Roehrig, T. Yu, and E. Krupinski, "Image quality control for digital mammographic systems: initial experience and outlook," J. Digit. Imaging **8**, 52–66 (1995).

[6] J. M. Boone, T. R. Fewell, and R. J. Jennings, "Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography," Med. Phys. **24**, 1863–1874 (1997).

[7] A. Karellas, I. Sechopoulos, I. Levis, A. C. Huber, and J. A. Pantazis, "Measurement of the x-ray spectra and tube potential in mammographic units with a self-calibrating compact cadmium zinc telluride (CZT) detector," Radiology **205**, 301 (1997).

[8] Title 21, United States Code of Federal Regulations, Part 900, 1998.

[9] U. Bick, M. L. Giger, R. A. Schmidt, R. M. Nishikawa, and K. Doi, "Density correction of peripheral breast tissue on digital mammograms," Radiographics **16**, 1403–1411 (1996).

[10] M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," IEEE Trans. Med. Imaging **17**, 510–517 (1998).

[11] S. S. Buchbinder, I. S. Leichter, P. N. Bamberger, B. Novak, R. Lederman, S. Fields, and D. J. Behar, "Analysis of clustered microcalcifications by using a single numeric classifier extracted from mammographic digital images," Acad. Radiol. **5**, 779–784 (1998).

[12] M. A. Anastasio, H. Yoshida, R. Nagel, R. M. Nishikawa, and K. Doi, "A genetic algorithm-based method for optimizing the performance of a computer-aided diagnosis scheme for detection of clustered microcalcifications in mammograms," Med. Phys. **25**, 1613–1620 (1998).

[13] D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," Med. Phys. **24**, 903–914 (1997).

[14] H. Yoshida, K. Doi, R. M. Nishikawa, M. L. Giger, and R. A. Schmidt, "An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms," Acad. Radiol. **3**, 621–627 (1996).

[15] M. Kallergi, G. M. Carney, and J. Gaviria, "Evaluating the performance of detection algorithms in digital mammography," Med. Phys. **26**, 267–275 (1999).

[16] L. T. Niklason *et al.*, "Digital tomosynthesis in breast imaging," Radiology **205**, 399–406 (1997).

[17] U. E. Ruttimann, R. A. J. Groenhuis, and R. L. Webber, "Restoration of digital multiplane tomosynthesis by a constrained iteration method," IEEE Trans. Med. Imaging **MI-3**, 141–148 (1984).

[18] D. P. Chakraborty and G. T. Barnes, "An energy sensitive cassette for dual-energy mammography," Med. Phys. **16**, 7–13 (1989).

[19] T. Asaga, S. Chiyasu, S. Mastuda, H. Mastuura, H. Kato, M. Ishida, and T. Komaki, "Breast imaging: Dual-energy projection radiography with digital radiography," Radiology **164**, 869–870 (1987).

[20] F. L. Flanagan, J. G. Murray, P. Gilligan, J. P. Stack, and J. T. Ennis, "Digital subtraction in Gd-DTPA enhanced imaging of the breast," Clin. Radiol. **50**, 848–854 (1995).

[21] M. B. Williams and L. L. Fajardo, "Digital mammography: Performance considerations and current detector designs," Acad. Radiol. **3**, 429–437 (1996).

[22] D. W. Holdsworth, R. K. Gerson, and A. Fenster, "A time-delay integration charge-coupled device camera for slot-scanned digital radiography," Med. Phys. **17**, 876–886 (1990).

[23] R. M. Nishikawa, G. E. Mawdsley, A. Fenster, and M. J. Yaffe, "Scanned-projection digital mammography," Med. Phys. **14**, 717–727 (1987).

[24] A. Karellas, L. J. Harris, H. Liu, M. A. Davis, and C. J. D'Orsi, "Charge-coupled device detector: Performance considerations and potential for small-field mammographic imaging applications," Med. Phys. **19**, 1015–1023 (1992).

[25] S. Hejazi and D. P. Trauernicht, "System considerations in CCD-based x-ray imaging for digital chest radiography and digital mammography," Med. Phys. **24**, 287–297 (1997).

[26] J. H. Siewerdsen, L. E. Antonuk, Y. El-Mohri, J. Yorkston, W. Huang, J. M. Boudry, and I. A. Cunningham, "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology," Med. Phys. **24**, 71–89 (1997).

[27] L. E. Antonuk *et al.*, "Performance evaluation of a large area, 97 μm pitch: Indirect detection active matrix flat-panel imager (AMFPI) for radiography and fluoroscopy," Radiology **209**, 357 (1998).

[28] L. E. Antonuk *et al.*, "Performance limits of high resolution large area active matrix flat-panel imagers (AMFPIs)," Radiology **209**, 581 (1998).

[29] L. E. Antonuk *et al.*, "Strategies to significantly enhance performance of active matrix flat-panel imagers (AMFPIs)," Radiology **209**, 358 (1998).

[30] H. M. Rougeot, B. Opsahl-Ong, D. E. Castleberry, C. E. Landberg, J. Q. Liu, and C. M. Kimme-Smith, "Performance evaluation of a flat-panel filmless full-field digital mammography system," Radiology **201**, 190 (1996).

[31] U. Neitzel, I. Maack, and S. Gunther-Kohfahl, "Image quality of a digital chest radiography system based on a selenium detector," Med. Phys. **21**, 509–516 (1994).

[32] W. Zhao and J. A. Rowlands, "Digital radiology using active matrix readout of amorphous selenium: Theoretical analysis of detective quantum efficiency," Med. Phys. **24**, 1819–1833 (1997).

[33] S. L. Garverick, L. Skrenes, and R. D. Baertsch, "A 32-channel charge readout IC for programmable, nonlinear quantization of multichannel detector data," IEEE J. Solid-State Circuits **30**, 533–541 (1995).

[34] H. Fujita, D. Y. Tsai, T. Itoh, K. Doi, J. Morishita, K. Ueda, and A. Ohtsuka, "A simple method for determining the modulation transfer function in digital radiography," IEEE Trans. Med. Imaging **11**, 34–39 (1992).

[35] J. T. Dobbins, D. L. Ergun, L. Rutz, D. A. Hinshaw, H. Blume, and D. C. Clark, "DQE(f) of four generations of computed radiography acquisition devices," Med. Phys. **22**, 1581–1593 (1995).

[36] J. T. Dobbins, "Effects of undersampling on the proper interpretation of modulation transfer function, noise power spectra, and noise equivalent quanta of digital imaging systems," Med. Phys. **22**, 171–181 (1995).

[37] M. L. Giger, K. Doi, and C. E. Metz, "Investigation of basic imaging properties in digital radiography. II. Noise Weiner spectrum," Med. Phys. **11**, 797–805 (1984).

[38] C. D. Bradford, W. W. Peppler, and J. T. Dobbins, "Performance characteristics of a Kodak computed radiography system," Med. Phys. **26**, 27–37 (1999).

[39] J. C. Dainty and R. Shaw, *Image Science* (Academic, New York, 1974).

[40] H. E. Johns and J. R. Cunningham, *The Physics of Radiology*, 4th ed. (Thomas, Springfield, IL, 1983).

[41] J. W. Byng, J. G. Mainprize, and M. J. Yaffe, "X-ray characterization of breast phantom materials," Phys. Med. Biol. **43**, 1367–1377 (1998).

[42] R. M. Nishikawa and M. J. Yaffe, "SNR properties of mammographic film-screen systems," Med. Phys. **12**, 32–39 (1985).

[43] P. Bunch, "The effects of reduced film granularity on mammographic image quality," Proc. SPIE **3032**, 302–317 (1997).

[44] L. T. Niklason *et al.*, "Improved detection of low contrast objects with full-field digital mammography versus film-screen mammography," Radiology **205**, 436 (1997).

[45] R. E. Hendrick *et al.*, "Low-contrast lesion detection: Comparison of

screen-film and full-field digital mammography," Radiology **205**, 274 (1997).
[46] A. Karellas *et al.*, "Evaluation of a full-field clinical prototype flat panel imager for digital mammography," Radiology **209**, 159 (1998).
[47] J. M. Lewin *et al.*, "Clinical evaluation of a full-field digital mammography prototype for cancer detection in a screening setting—work in progress," Radiology **209**, 238 (1998).
[48] L. Moss, C. J. D'Orsi, A. Karellas, E. Hendrick, J. Lewin, and G. Sisney, "Initial experience with a high resolution full field digital mammographic system," J. Digit. Imaging **11**, 110 (1998).
[49] R. H. Moore *et al.*, "Initial clinical experience with full-field digital mammography," Radiology **205**, 274 (1997).